

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:51:59 ; Search time 5280.09 Seconds  
(without alignments)  
11342.728 Million cell updates/sec

Title: US-09-831-804-1\_COPY\_720\_1955

Perfect score: 1236  
Sequence: 1 atgagtcgaagtcgacgaac.....aaacatcagtcattcttcga 1236

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl!:\*  
1: gb Da:\*  
2: gb Htg:\*  
3: gb In:\*  
4: gb Om:\*  
5: gb Ov:\*  
6: gb Pat:\*  
7: gb Ph:\*  
8: gb Pl:\*  
9: gb Pr:\*  
10: gb Ro:\*  
11: gb Sts:\*  
12: gb Sy:\*  
13: gb Un:\*  
14: gb Vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1236	100.0	1239	6	BD274350	BD274350 Candida a
2	1236	100.0	2060	6	BD274349	BD274349 Candida a
3	1232.8	99.7	1239	6	AX488193	AX488193 Sequence
4	581.8	47.1	627	6	ARS50053	ARS50053 Sequence
5	438.4	35.5	110000	8	CR382135_06	Continuation (7 of
6	391	31.6	462	6	ARS50055	ARS50055 Sequence
7	228	18.4	110000	8	CR382122_03	Continuation (4 of
8	227.2	18.4	92211	8	CR380953_09	Continuation (10 o
9	217.4	17.6	1560	8	YSCNOFEAT	M50638 Saccharomyc
10	217.4	17.6	1739	8	YSCPTFIIA	M80611 Saccharomyc
11	217.4	17.6	37497	8	YSCP9677	U55841 Saccharomyc
12	199.2	16.1	2133	8	YSCRP026A	M33924 S.cerevisia
13	193	15.6	797	11	CNS061NG	AL400514 T7 end of
14	120.6	9.8	110000	8	AE016818_01	Continuation (2 of
15	101	8.2	169063	9	AC138126	AC138126 Homo sapi
16	101	8.2	177299	9	AC073544	AC073544 Homo sapi
17	99.4	8.0	2132	9	AX714042	AX714042 Sequence
18	99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
19	99.4	8.0	2196	6	CQ726027	CQ726027 Sequence

C 20	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
C 21	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
C 22	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
C 23	97.8	7.9	1203	9	D70831	D70831 Homo sapien
C 24	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
C 25	97.2	7.9	115955	9	AC011494	AC011494 Homo sapi
C 26	96.6	7.8	2320	6	AR270491	AR270491 Sequence
C 27	96.6	7.8	2330	6	HSU5376	US3736 Human repre
C 28	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
C 29	96.6	7.8	106811	9	HS56705	CR556705 Mus muscu
C 30	95	7.7	981	9	HS42P9	X78932 H.sapiens H
C 31	94.6	7.7	186233	9	AC092329	AC092329 Homo sapi
C 32	94.6	7.7	189317	2	AC024483	AC024483 Homo sapi
C 33	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
C 34	93.6	7.6	1389	9	M27879	M27879 Homo sapien
C 35	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
C 36	93.4	7.6	199585	10	AC124426	AC124426 Mus muscu
C 37	92.4	7.5	3516	6	CQ726030	CQ726030 Sequence
C 38	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
C 39	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
C 40	92	7.4	2873	9	HUM6PLK	M55432 Human Kruep
C 41	92	7.4	2873	11	G28705	G28705 EMS3976 Er
C 42	92	7.4	2873	11	HUMSW51269	G18281 human chrom
C 43	92	7.4	156835	9	AC010620	AC010620 Homo sapi
C 44	92	7.4	191082	2	AC141066	AC141066 Homo sapi
C 45	92	7.4	203396	9	AC073210	AC073210 Homo sapi

## ALIGNMENTS

RESULT 1	BD274350	1239 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD274350				
DEFINITION	Candida albicans tffIIIGene (CatfIIIA) and the coded CATfIIIA				
ACCESSION	BD274350				
VERSION	BD274350.1	GI:33084118			
KEYWORDS	JP 2002531068-A/2.				
SOURCE	Candida albicans				
ORGANISM	Candida albicans				
REFERENCE	1 (bases 1 to 1239)				
AUTHORS	Pallier, F.B., Camier, S. and Sentenac, A.				
TITLE	Candida albicans tffIIIGene (CatfIIIA) and the coded CATfIIIA				
JOURNAL	Patent: JP 2002531068-A 2 24-SEP-2002;				
COMMENT	AVENTIS PHARMA SA				
OS	Candida albicans				
PN	JP 2002531068-A/2				
PD	24-SEP-2002				
PF	09-NOV-1999	JP 2000581204			
PR	10-NOV-1998	FR 98/14147			
PI	FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC	PC			
C12N15/09, C12N15/09, A61K39/00, A61K5/00, A61P31/10, C07K14/40, PC					
C07K16/14, PC					
PC C12N1/19, C12N1/21, C12P21/02, C12Q1/02, G01N33/15, G01N33/50, PC					
PC G01N33/53, PC					
PC G01N33/566, G01N33/569, C12N15/09, C12R1.725, C12N15/00, C12R1.725					
PC 00, PC					
CC (C12N15/00, C12R1.725)					
CC Candida albicans tffIIIGene (CatfIIIA) and the coded CATfIIIA					
CC CC					
CC Key					
CC Key					
FT CDS					
Location/Qualifiers					
1..1239					
/organism="Candida albicans"					
/mol_type="genomic DNA"					
/db_xref="taxon:5476"					

Query Match 100.0%; Score 1236; DB 6; Length 1239;



QY 181 GATTGATTAAGCATTTTTCAGAAAATCAATTTGGAAACATAATGTATCAATCTCC 240  
|  
|  
|  
Db 900 GATTGATTAAGCATTTTTCAGAAAATCAATTTGGAAACATAATGTATCAATCTCC 959  
|  
|  
|  
QY 241 GAAAAAAACCATTTCCATTGTTCAGTGTGTGTGTAAAGGGTTAATCTCGACAACCTTG 300  
|  
|  
|  
|  
Db 960 GAAAAAAACCATTTCCATTGTTCAGTGTGTGTAAAGGGTTAATCTCGACAACCTTG 1019  
|  
|  
|  
QY 301 AAAAGCATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAA 360  
|  
|  
|  
|  
Db 1020 AAAAGCATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAA 1079  
|  
|  
|  
QY 361 GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTA 420  
|  
|  
|  
|  
Db 1080 GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTA 1139  
|  
|  
|  
QY 421 ACGGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTTAGACAACATTAATTA 480  
|  
|  
|  
|  
Db 1140 ACGGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTTAGACAACATTAATTA 1199  
|  
|  
|  
QY 481 AAACATCATGTGTGATCTCTGCTTATCAATGTATCATCTGTTGTTTAAATTTTC 540  
|  
|  
|  
|  
Db 1200 AAACATCATGTGTGATCTCTGCTTATCAATGTATCATCTGTTGTTTAAATTTTC 1259  
|  
|  
|  
QY 541 CAAACTTGTCTAGTATTACATTTTCAATATATAAACAATGATCCAAACTTAATGCTCT 600  
|  
|  
|  
|  
Db 1260 CAAACTTGTCTAGTATTACATTTTCAATATATAAACAATGATCCAAACTTAATGCTCT 1319  
|  
|  
|  
QY 601 AAAATGTGTAAAGGTGTGTGGAAAAAGTTTATCTTCAATATGTTAATGATCATGAT 660  
|  
|  
|  
|  
Db 1320 AAAATGTGTAAAGGTGTGTGGAAAAAGTTTATCTTCAATATGTTAATGATCATGAT 1379  
|  
|  
|  
QY 661 GATTTCACATGATCAAAATATGACCTTGTATATGTATGTATGTGGGAAATTTGCAAG 720  
|  
|  
|  
|  
Db 1380 GATTTCACATGATCAAAATATGACCTTGTATATGTATGTATGTGGGAAATTTGCAAG 1439  
|  
|  
|  
QY 721 AAAAATGAATTAAGTGAACATTTAATATCTTCCATGATGTAAATCTCTGATGATTA 780  
|  
|  
|  
|  
Db 1440 AAAAATGAATTAAGTGAACATTTAATATCTTCCATGATGTAAATCTCTGATGATTA 1499  
|  
|  
|  
QY 781 TTTAAAGGAACTGAAGTGAAGAAAAATTAGAACCTTATTAATCAAGATGAAATTAAT 840  
|  
|  
|  
|  
Db 1500 TTTAAAGGAACTGAAGTGAAGAAAAATTAGAACCTTATTAATCAAGATGAAATTAAT 1559  
|  
|  
|  
QY 841 AATTTGATGATTAAGAAACAGAGAAATTTAAAGTGAAGAAAGTGAAGAAAGTGAAGAA 900  
|  
|  
|  
|  
Db 1560 AATTTGATGATTAAGAAACAGAGAAATTTAAAGTGAAGAAAGTGAAGAAAGTGAAGAA 1619  
|  
|  
|  
QY 901 GATAGCTAGATGAAAAAGAGATGATTAAGATCAGATCAATGTGAGCTCAAGATCA 960  
|  
|  
|  
|  
Db 1620 GATAGCTAGATGAAAAAGAGATGATTAAGATCAGATCAATGTGAGCTCAAGATCA 1679  
|  
|  
|  
QY 961 ATAAATCATTTTACCTGCTTCTTGGAGGTTCAAGAGTGTCTTAACTTATCTGAAT 1020  
|  
|  
|  
|  
Db 1680 ATAAATCATTTTACCTGCTTCTTGGAGGTTCAAGAGTGTCTTAACTTATCTGAAT 1739  
|  
|  
|  
QY 1021 AGTGGGAAAGATCAATTTGCTTGAAGATTAATTTGTATGATGTTTCTAAGAAATAT 1080  
|  
|  
|  
|  
Db 1740 AGTGGGAAAGATCAATTTGCTTGAAGATTAATTTGTATGATGTTTCTAAGAAATAT 1799  
|  
|  
|  
QY 1081 GATTTCATGATGAAATGATGATGATTAATTAACAAGAAATTTGATGATCTTA 1140  
|  
|  
|  
|  
Db 1800 GATTTCATGATGAAATGATGATGATTAATTAACAAGAAATTTGATGATCTTA 1859  
|  
|  
|  
QY 1141 AATAGTATGAAAAAGAGAAATCTCAGAAAGTGAACCATTTGTTAAAAAGCAGAGATG 1200  
|  
|  
|  
|  
Db 1860 AATAGTATGAAAAAGAGAAATCTCAGAAAGTGAACCATTTGTTAAAAAGCAGAGATG 1236  
|  
|  
|  
QY 1201 GATTTCATGATGAAATGATGATGATTTCTCGA 1236  
|  
|  
|  
|  
Db 1920 GATTTCATGATGAAATGATGATGATTTCTCGA 1955

RESULT 3  
AX489193 1239 bp DNA linear PAT 16-AUG-2002  
LOCUS  
DEFINITION Sequence 6493 from Patent WO02053728.  
ACCESSION AX489193  
VERSION AX489193.1 GI:22323205  
KEYWORDS  
SOURCE  
ORGANISM  
Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Microsporid Saccharomycetales; Candida.  
REFERENCE  
1  
AUTHORS Roemer, T., Jiang, B., Boone, C., Bussey, H., and Ohlsen, K. L.  
TITLE Gene disruption methodologies for drug target discovery  
JOURNAL Patent: WO 02053728-A 6493 11-JUL-2002;  
Elitza Pharmaceuticals, Inc. (US)  
FEATURES  
source  
1..1239  
Location/Qualifiers  
/organism="Candida albicans"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5476"  
ORIGIN  
Query Match 99.7%; Score 1232.8; DB 6; Length 1239;  
Best Local Similarity 99.8%; Pred. No. 4.8e-189;  
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGAGTGAAGATGACGAAACCAATGATATCATCTTTAATATCTTCTTCTTATCA 60  
|  
|  
|  
|  
Db 1 ATGAGTGAAGATGACGAAACCAATGATATCATCTTTAATATCTTCTTCTTATCA 60  
|  
|  
|  
QY 61 CGTCCCAAAAAGATATTTTGACATATGAAAGGGTGTATTAAGCCTATATGACCATCA 120  
|  
|  
|  
|  
Db 61 CGTCCCAAAAAGATATTTTGACATATGAAAGGGTGTATTAAGCCTATATGACCATCA 120  
|  
|  
|  
QY 121 TTAATGAGCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAC 180  
|  
|  
|  
|  
Db 121 TTAATGAGCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAC 180  
|  
|  
|  
QY 181 GATTGATTAAGCATTTTTCAGAAAATCACTTTGGAAACATATTTGTATCAATCTCC 240  
|  
|  
|  
|  
Db 181 GATTGATTAAGCATTTTTCAGAAAATCACTTTGGAAACATATTTGTATCAATCTCC 240  
|  
|  
|  
QY 241 GAAAAAAACCATTTCCATTGTTCAGTGTGTGTAAAGGGTTAATCTCGACAACCTTG 300  
|  
|  
|  
|  
Db 241 GAAAAAAACCATTTCCATTGTTCAGTGTGTGTAAAGGGTTAATCTCGACAACCTTG 300  
|  
|  
|  
QY 301 AAAAGCATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAA 360  
|  
|  
|  
|  
Db 301 AAAAGCATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAA 360  
|  
|  
|  
QY 361 GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTA 420  
|  
|  
|  
|  
Db 361 GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTA 420  
|  
|  
|  
QY 421 ACGGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTTAGACAACATTAATTA 480  
|  
|  
|  
|  
Db 421 ACGGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTTAGACAACATTAATTA 480  
|  
|  
|  
QY 481 AAACATCATGTGTGATCTCTGCTTATCAATGTATCATCTGTTGTTTAAATTTTC 540  
|  
|  
|  
|  
Db 481 AAACATCATGTGTGATCTCTGCTTATCAATGTATCATCTGTTGTTTAAATTTTC 540  
|  
|  
|  
QY 541 CAAACTTGTCTAGTATTACATTTTCAATATTAACAACATGATCCAAACTTAATGCTCT 600  
|  
|  
|  
|  
Db 541 CAAACTTGTCTAGTATTACATTTTCAATATTAACAACATGATCCAAACTTAATGCTCT 600  
|  
|  
|  
QY 601 AAAATGTGTAAAGGTGTGTGGAAAAAGTTTATCTTCAATATGTTAATGATCATGAT 660  
|  
|  
|  
|  
Db 601 AAAATGTGTAAAGGTGTGTGGAAAAAGTTTATCTTCAATATGTTAATGATCATGAT 660  
|  
|  
|  
QY 661 GATTTCATGATCAAAATATGACCTTGTATATTTGTATGATGTGGGAAATTTGCAAG 720  
|  
|  
|  
|

Db 661 GATTTCACATGATCAAAAATATGACTTGATTATTTGTGATGGGAAAATTTCGAAG 720  
Qy 721 AAAAATGAATTAAGTTGAACATTATTAATCTTCATGATGTAATATCCCTGATGATTA 780  
Db 721 AAAAATGAATTAAGTTGAACATTATTAATCTTCATGATGTAATATCCCTGATGATTA 780  
Qy 781 TTTAAGAAACCTGAAGTGAATAAATTAGAACCTTATTAGATCAAGATCGAAATTTAAAT 840  
Db 781 TTTAAGAAACCTGAAGTGAATAAATTAGAACCTTATTAGATCAAGATCGAAATTTAAAT 840  
Qy 841 AATTTCATGAATTTAGAAACAGAGAAATTTAAAGTGAAGAAAGTGAAGAAATTTGAAGA 900  
Db 841 AATTTCATGAATTTAGAAACAGAGAAATTTAAAGTGAAGAAAGTGAAGAAATTTGAAGA 900  
Qy 901 GATAGCTAGATGAAAAAGAAAGATGATGTTAGATCAAGCTCAATGTCAGCTCAAAATCA 960  
Db 901 GATAGCTAGATGAAAAAGAAAGATGATGTTAGATCAAGCTCAATGTCAGCTCAAAATCA 960  
Qy 961 ATTTAAATCATTTAATCTGCTTCTTTGGAAGTTCAAAGAGTTTCTAACTTATTCGAAT 1020  
Db 961 ATTTAAATCATTTAATCTGCTTCTTTGGAAGTTCAAAGAGTTTCTAACTTATTCGAAT 1020  
Qy 1021 AGTGGGAAGAAAGATCAATTTGCTTAAGAAATTTGATGAATGTTTCTTAAGAAATAT 1080  
Db 1021 AGTGGGAAGAAAGATCAATTTGCTTAAGAAATTTGATGAATGTTTCTTAAGAAATAT 1080  
Qy 1081 GATTTCAGTCGACATTGGAATATGATGATGATTAATTTACAAAGAAATGATGATCTTAA 1140  
Db 1081 GATTTCAGTCGACATTGGAATATGATGATGATTAATTTACAAAGAAATGATGATCTTAA 1140  
Qy 1141 AATAGTATAGAAAAAGAAAGAACTCCAGAAAGTGAACCAATGGTTAAAAAGCCAGAGATG 1200  
Db 1141 AATAGTATAGAAAAAGAAAGAACTCCAGAAAGTGAACCAATGGTTAAAAAGCCAGAGATG 1200  
Qy 1201 GATTTCAGCAATGAATGATGATGATTTCTCGA 1236  
Db 1201 GATTTCAGCAATGAATGATGATGATTTCTCGA 1236

RESULT 4  
AR550053  
LOCUS AR550053 627 bp DNA 1linear PAT 08-OCT-2004  
DEFINITION Sequence 5184 from patent US 6747137.  
ACCESSION AR550053  
VERSION AR550053.1 GI:53943228  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 627)  
AUTHORS Weinstein, K.G. and Bush, D.  
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics  
and therapeutics  
JOURNAL Patent: US 6747137-A 5184 08-JUN-2004;  
FEATURES  
source  
1. 627  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 47.1%; Score 581.8; DB 6; Length 627;  
Best Local Similarity 97.4%; Pred. No. 3.6e-84;  
Matches 605; Conservative 0; Mismatches 7; Indels 9; Gaps 1;  
Qy 616 TGTGTTGGGAAAAAGGTTTATCTTCATATGTTAAGTCATGATGTTCTACATGATC 675  
Db 13 TGTGTTGGGAAAAAGGTTTATCTTCATATGTTAAGTCATGATGTTTTCACATGATC 72  
Qy 676 AAAATATGACTGTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 735  
Db 73 AAAATATGACTGTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 132  
Qy 736 GAACATTATATATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 795

Db 133 GAACATTATATATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192  
Qy 796 GTGAATAATTAAGAACCTATTTAGATCAAGATCGAAATTTAAATTTGATGATTA 855  
Db 193 GTGAATAATTAAGAACCTATTTAGATCAAGATCGAAATTTAAATTTGATGATTA 252  
Qy 856 GAACAGAAATTTAAAGTGGAAAGATGAAAGATGAAAGATGATGATGATGATGATGATG 915  
Db 253 GAACAGAAATTTAAAGTGG-----GAAGAAATGAAAGATGATGATGATGATGATG 303  
Qy 916 AAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 975  
Db 304 AAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363  
Qy 976 GCTTCCTTGGAAAGTTCAAAGAGTTTCTAACTTATTTGATGATGATGATGATGATGATG 1035  
Db 364 GCTTCCTTGGAAAGTTCAAAGAGTTTCTAACTTATTTGATGATGATGATGATGATGATG 423  
Qy 1036 AATTGCTTAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095  
Db 424 AATTGCTTAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 483  
Qy 1096 TTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155  
Db 484 TTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543  
Qy 1156 GAAGAAATCCCAAGAGTGAACCATTTGTTAAAAAGCCAGATGATGATGATGATGATG 1215  
Db 544 GAAGAAATCCCAAGAGTGAACCATTTGTTAAAAAGCCAGATGATGATGATGATGATG 603  
Qy 1216 GAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1236  
Db 604 GAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 624

RESULT 5  
CR382135\_06/c  
WPCOMMENT  
Sequence split into 16 fragments LOCUS CR382135 Accession CR382135  
Fragment Name Begin End  
CR382135\_00 1 110000  
CR382135\_01 100001 210000  
CR382135\_02 200001 310000  
CR382135\_03 300001 410000  
CR382135\_04 400001 510000  
CR382135\_05 500001 610000  
CR382135\_06 600001 710000  
CR382135\_07 700001 810000  
CR382135\_08 800001 910000  
CR382135\_09 900001 1010000  
CR382135\_10 1000001 1110000  
CR382135\_11 1100001 1210000  
CR382135\_12 1200001 1310000  
CR382135\_13 1300001 1410000  
CR382135\_14 1400001 1510000  
CR382135\_15 1500001 1592360  
Continuation (7 of 16) of CR382135 from base 600001 (CR382135 Debaryomyces hanseii chro  
Query Match 35.5%; Score 438.4; DB 8; Length 110000;  
Best Local Similarity 63.4%; Pred. No. 1.3e-61;  
Matches 724; Conservative 0; Mismatches 406; Indels 12; Gaps 3;  
Qy 33 ATCTTAATATCTTCTCTCTTCATCACTGCTCCCAAAAGTATATTGACATATGAAGG 92  
Db 20471 ATCTTAATATCTTCTCTCTTCATCACTGCTCCCAAAAGTATATTGACATATGAAGG 20412  
Qy 93 GTGTATTAAGGCTTAATTCGACATCATTTATTAGCAACATTTAAGAACCCAGATGA 152  
Db 20411 GTGTATTAAGGCTTAATTCGACATCATTTATTAGCAACATTTAAGAACCCAGATGA 20352  
Qy 153 TGATGACCGTATTAATGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 212



Db 20351 TGAAGAAGCCCTTAAATGCTCGAAGAGCTGATGCGAAGTCAATCTTCTTAAGAAAGTCACA 20292  
Qy 213 TTGGAAACACATATGATC-----ACATTCGAAAAAAACCATTCATGTTGAGT 266  
Db 20291 TTTTCAAGACATTTGCTTTCTCTCATGAAAGTCAAGAAAGTAAACCTTCCAGTGTCAGC 20232  
Qy 267 GTGTGTAAAGGGGTTAAATCTCGACAAACCTTGAAAGACATGAAATCAACCCATACAAA 326  
Db 20231 TTGGGCAAAAGGTGTAATACCTTACAAACATTTAAAGACACGAATTAACATACATAAA 20172  
Qy 327 GTCAATTAATGTCATTTGAAATTTGCAAGAGCATTTTATTAACATCAATCTTTAG 386  
Db 20171 ATCAATTAATTTGTCATTTGGGGCTGTAAATGATCAATTTATTAAGATCAGTCACAG 20112  
Qy 387 ACATCATATATTAATCTGTCATGAAAGAAACATTAAGCTTAACATGATATAAGTTT 446  
Db 20111 ACACCATACGCTATCGGTTCTATGAAAAAAATTTATTAAGCAATAGTAAATATCGTT 20052  
Qy 447 CACTGACCTTCAAAATTAAGCAACAACTAAATTAACATCATGATGATCTCCTGCTTA 506  
Db 20051 TTAACAGACCGTATGATTTGGACAGCAACAAATTAATATCAGTACCTGCGGCTTA 19992  
Qy 507 TCAATGTGATCAGCTGCTGTTTAAAAATTTCCAACTTGTCAGTATTAACAATTTCA 566  
Db 19991 TCAATGTGATCAGCTGCTGTTTAAAAATTTATGACGTCGCTTTACATTTGCA 19932  
Qy 567 TATTAACCACTGATCCAAACCTTAATGTCCTTAATGTGTAAAGTTGTGTGGAA 626  
Db 19931 TATTAACCACTGAGACCCAAAGATTAAGTCTATATGTGGAAAGTGTGCGGTAA 19872  
Qy 627 AAAAGTTTATCTTCAATATGTTAAGTCATGATGATTTTACATGATCAAAATATGAC 686  
Db 19871 GAAAGGCTTCGTCACACATGATATTTACGATGAAAGAAAGTATGCAATTAATGAA 19812  
Qy 687 TTGTGATTTATGTGATGTGGGAAATTTGCAAGAAATGAAATTAAGTGAACATTA 746  
Db 19811 TTGTGATTTATGTGATGTGGAAATTTGCAAGAAATGAAATTAAGTGAACATTA 19752  
Qy 747 TATTTTCCATGATGATTAATATCCCTGATGATTTTAAAGAACTGAAGTGAAGAAAT 806  
Db 19751 CACTTATCAGACAGAAATCTTCTGATGATTTATGAAACCGATCGAACGAGAACGTT 19692  
Qy 807 AGAAGACCTATTA---GATCAAGATCGAAATTAATTAATTTGATGATTAAGAAACAGA 863  
Db 19691 AGATTAACCTTAAAGCAGAAAGACTTACCAATATGATTTTAATCTTTAGAGATTT 19632  
Qy 864 GAAATTAAGTGAAGAGATGAAGAGATGAAGAAAGATGCTAG---ATGAAAAAG 920  
Db 19631 ACAATCAAAAGGTTTGTGAAGTACCGTCGATGAAGAGAAAGAGATGATTCAT 19572  
Qy 921 AAGTATGTTAGATCAGACTCAATGCTCAAGATCAATTAATCATTTACTGCTTC 980  
Db 19571 GTCCAAATGTTGCAAGCTTCACATTAATCTTTGAATTCGTTGAATTTAACTTTGGAATC 19512  
Qy 981 TTGGAAGGTTCAAAGAGTGTTCGAACTTATCTGAATGATGAGGAAGATCAATTTG 1040  
Db 19511 TGGAAAGCATCTTATGTAGTCTTATTTCTGAATTAATTTATGAAAGGAAAGTTCCATG 19452  
Qy 1041 TCCTAAGAAATATGTGATAGATGTTTCTAGAAATATGATTAAGTCGACATTTGAA 1100  
Db 19451 TCCGAAAAAGAAATGTGATTAATGTTACGCGGATTAATGATCTTGAAGACATTTAA 19392  
Qy 1101 ATGCGATGATATTAATTTCAAGAAATTTGATCTTAAATGATAGAAAAAGAA 1160  
Db 19391 GTGCGATGAATCTCAATTAAGAAATTTGAAGCGTTTAAAAAGTCTTGAAGTCGAAGA 19332  
Qy 1161 AA 1162  
Db 19331 AA 19330

RESULT 6  
ARS50055

LOCUS ARS50055 462 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 5186 from patent US 6747137.  
ACCESSION ARS50055  
VERSION ARS50055.1 GI:53943230  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 462)  
AUTHORS Weinstein, K.G. and Bush, D.  
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics  
and therapeutics  
JOURNAL Patent: US 6747137-A 5186 08-JUN-2004;  
FEATURES  
source Location/Qualifiers  
1..462  
/organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

Query Match 31.6%; Score 391; DB 6; Length 462;

Best Local Similarity 98.5%; Pred. No. 2e-53; Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 ATGAGTGAAGGTAGCAGAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCA 60  
Db 52 ATGAGTGAAGGTAGCAGAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCA 111  
Qy 61 CGTCCCAAAAGTATATTTGACATATGAAAGGTGTGATTAAGCCTTAATTCAGACATCA 120  
Db 112 CGTCCCAAAAGTATATTTGACATATGAAAGGTGTGATTAAGCCTTAATTCAGACATCA 171  
Qy 121 TTAATGAGCAACATTTAAGACCCACGTAATGATGACCGTATTAATGATGACGTGAC 180  
Db 172 TTAATGAGCAACATTTAAGACCCACGTAATGATGACCGTATTAATGATGACGTGAC 231  
Qy 181 GATTGTGATTAAGCATTTTTCAGAAATCAATTTGGAAACATATTTGATCATCTTC 240  
Db 232 GATTGTGATTAAGCATTTTTCAGAAATCAATTTGGAAACATATTTGATCATCTTC 291  
Qy 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTTCGACAACTTTG 300  
Db 292 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTTCGACAACTTTG 351  
Qy 301 AAAAGCATGAATTCACCCATCAAGTCAATTAATGTCATTTGAAATTTGCAAGAA 360  
Db 352 AAAAGCATGAATTCACCCATCAAGTCAATTAATGTCATTTGAAATTTGCAAGAA 411  
Qy 361 GCATTTTAT-AAACATCAATCTTTAAGACATCATATTAATCTGTTTCATGA 410  
Db 412 ACATTTTATTAACATCAATCTTTAAGACATCATATTAATCTGTTTCATGA 462

## RESULT 7

CR382122 03/c

WPCOMMENT

Sequence split into 14 fragments LOCUS CR382122 Accession CR382122

Fragment Name	Begin	End
CR382122_00	1	110000
CR382122_01	100001	210000
CR382122_02	200001	310000
CR382122_03	300001	410000
CR382122_04	400001	510000
CR382122_05	500001	610000
CR382122_06	600001	710000
CR382122_07	700001	810000
CR382122_08	800001	910000
CR382122_09	900001	1010000
CR382122_10	1000001	1110000
CR382122_11	1100001	1210000
CR382122_12	1200001	1310000
CR382122_13	1300001	1320834

Continuation (4 of 14) of CR382122 from base 300001 (CR382122 Kluyveromyces fragilis strain)

Query Match 228; 18.4%; Score 228; DB 8; Length 110000;  
 Best Local Similarity 58.1%; Pred. No. 9.6e-28;  
 Matches 443; Conservative 0; Mismatches 310; Indels 9; Gaps 2;

QY	15150	101	15090	QY	161	15033	QY	221	14973	QY	281	14913	QY	341	14853	QY	401	14793	QY	461	14733	QY	515	14673	QY	575	14613	QY	635	14553	QY	695	14493	QY	755	14433
TATCTTCTCTCTCATCATCGTCCCAAAAAGTATATTGTGACATATGAGGGTGTGATA	TGTGAGTTTCATTCGTCACAGGGTTCTAAAGTATATTTCTGGATTATCAAGGTGTATCA	AAGCTATTAATCGACATCATTTATTAGACAACATTTAAGAAACCAAGTATGATCGAC	AATGCTTCACAAGACCGAGTTTGCTTCACGAGACATCAAGCTAC--CGTGATCATGAAA	CGTAAATATGACAGTGGACGATGTGTATTAAGCAATTTTCAAGAAATCATCAATTTGAAA	TCMAACCGTTTAATATGCGACCTGTGGCAAGAAATTTGCGAAAAATATGACATCTGAATA	CACATATTGTATCATCATTCGCAAAAAAACCATTTCCATTGTTCAGTGTGTGTAAAGGG	GACACATGTTCTCTCATATCAGATGATAAACCCTTTCACATGTTCTTATATGTGGAAAGTG	TTAATTCTCGACACACTTGAAGAAGCATGAAATCACCCATTAACAAAGTCATTTAAATGTA	TCACAACGACGACACAGTTAAAGACATGAAATTAACATACGAAGCTCTTCCACTGTG	CATTGAAAATATGTGCAAGAGCATTTTATAACATCAATCTTTTAAGACATATATATAT	CATATGAGGATGTATACGAATCATTTCTACAAACATCCAACTCCGATCGCATATCTTGT	CTGTTTCATGAAAAAACAATTAAACGTGTAAACATGATATATAAAGTTTCACTGCACCTTCAA	CGTTTCAGAAAAAGAAACTTACCTGCTCCCTCACTGTATACAGACGTTCCAAAGGCTTATC	AAATTGCAACAACATTAATTAAACATCATGTGTGATCTCTGTC-----TTATCATGTG	GATTGAAGAAATCATATAGATTAAGCATCAACACCAAGATCTACCGGAATGTACCAATGTG	ATCATCCGTGTGTTTAAATAATTTCCAAACCTGTGCTAGTATTCATTTGATTAATAAC	ACTTCTTCAATGACGACGATTTTCCACTTGCTCTTCTTTCACACAAACATTAATAAC	AACTGCATCCAAAACCTTAATATGTCCTTAATGTGTAAAGTGTGTGGAAAAAGGTT	AATGCAATCCGAATTAACATATGCTCTATTTGTGGAAACCTTGTGTGTGTAATCAGGTT	TATCTTCACATATGTTAAAGTATGATGATGATTTCTACCATGATCAAAATATGACTTGTGATT	TACGGAATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	ATTGGAATGTGGGAAAATTTGCAAAAGAAATAATGAATTAATGTTGAACATTATATATCTTCC	CATGTCACGATATCATCATTTTCCGAAAGAAACCTCAGCTAGTACACATTAATGAAAGAACTC	ATGATGTAATATATCCCTGATGATTTATTTAAAGAAACTGAAG	ACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT											

RESULT 8  
 CR380953\_09/c  
 WPCOMMENT  
 Sequence split into 10 fragments LOCUS CR380953 Accession CR380953  
 Fragment Name Begin End  
 CR380953\_00 1 110000  
 CR380953\_01 10001 21000  
 CR380953\_02 20001 31000  
 CR380953\_03 30001 41000  
 CR380953\_04 40001 51000  
 CR380953\_05 50001 61000  
 CR380953\_06 60001 71000  
 CR380953\_07 70001 81000  
 CR380953\_08 80001 91000  
 CR380953\_09 90001 99221  
 Continuation (10 of 10) of CR380953 from base 90001 (CR380953 Candida glabrata strain C

Query Match	Best Local Similarity	18.44%	Score 227.2	DB 8	Length 92211
Matches 4477	Conservative 0	Mismatches 318	Indels 9	Gaps 2	
QY	44	CTTCTTCTTCATCAGCTCCCAAAAGTAAATTTTGACATATGAAAGGTGTGATAAAG	103		
Db	60262	CATCTGATCCAGTGTAAAGGTGAAAGACATATCTTTGTAATTCGATATTTGACAAAG	60203		
QY	104	CCCTATATGACCATCATTTTATAGACCAATTTAAGAACCCACGATATGATCCACCT	163		
Db	60202	CCTTTAAAGCCGCTCTTGCTACAGAAACACAGAAATC--AGTCATCTTGGAGGA	60148		
QY	164	ATAAATGACAGTGCAGATTTGTATTAAGCATTTTTCAGAAAATCACATTTGAAAAC	223		
Db	60145	AACCTTGGAAATGATATTCATATGGAAGTTCATTTCTAATAAAGATCCACTTAGAGAGC	60086		
QY	224	ATAATGTATCACATTCGAAAAAAACATTCATCTGTTCACTGTGTGTAAAGGGGTAA	283		
Db	60085	ACTTATACACACACAGATGAAAGCCGTTTATTTGTTCAATTTGTGGAAAGGGGCTTA	60028		
QY	284	ATTCTGACAAACCTTGAAAGAATGAAATACCAATCCATCAAGATCATTTAATGACAT	343		
Db	60025	TTACTAGGACAACTGAAACGACATGAGGTATCTACACAAATGTTCAATTTGGAAT	59966		
QY	344	TTGAAAATTTGTCAAGAGCATTTTATTAACATCAATCTTTAAGACATCATATATATCTG	403		
Db	59965	ATGAAAGTTGTATATGAGAGTTTCTACAGACCCCTCAATTTAAGGGGTCTATTTTAGCG	59906		
QY	404	TTCTATGAAAAACATTAACGTGTAAACATATATTAAGTTTCACTGACCTTCAAAAT	463		
Db	59905	TTCTATTACAAAGTCTAAATATGTCAGAAATGCAACAAAGTTTTCAAAGACCTTACAGC	59846		
QY	464	TAGACAAACATTAATTTAAACATCAT-----GGTAGCTCTCGTCTTCAATGTGATC	517		
Db	59845	TCAAAATTCATATGCCAAGACCAATATTCAGATTTGTCAATGCTTATCAATGTACCT	59786		
QY	518	ATCCTGTTCTTTTAAAAATTTTCCAAACTTGTGCTATTTAGCAATTTTCATATTAACAC	577		
Db	59785	TCAGTGTGTTCTTAAAGAGTTTCAAAACATGGTCTGCGTTAAGATTACATGTTAAAAAG	59726		
QY	578	TGCATCCAAACCTTAATATGCTTAAATGTGTATAGAGTTGTGTGGAAAAAGTTTAT	637		
Db	59725	ATCATCCGAAGTTTAAATGCCCTATATGACATTAACCCCTGTGTGGGGAAGCGTTTAA	59666		
QY	638	CTTCACATATATTAAGCATGATGATTTTACACATGATCAAAATATGAGCACTTGATATAT	697		
Db	59665	ATATGACATGAAAAATCCAGATGAAAACTTGTTCAGAAATTTGGAAGTGTATATAT	59606		
QY	698	GTGATGTGGGAAATTTGCAGAAAAATGAATTAAGTGAACATTAATATCTTCCATG	757		
Db	59605	GTATATGATCAGTCATTTGCTTAAGAAATTTGAACCTATTTGATCATTAACACACATGCA	59546		
QY	758	ATGTATATATTCCTGATGATTTTATTAAGAAACATGAAGTGAAAAAATTTAGGA	811		
Db	59545	GTGAAGAAATCCCGGCTATCTATTGAACAGAAAGTATGCTTATGTTGAAA	59492		

common subunit RPB6 are divergently transcribed in *Saccharomyces cerevisiae*  
Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)  
MEDLINE 92237295  
PUBMED 1570325  
COMMENT Original source text: *Saccharomyces cerevisiae* (strain S288C) DNA.  
FEATURES  
source  
1. .1560  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="S288C"  
/db\_xref="taxon:4932"

ORIGIN chromosome 14.

Query Match 17.6%; Score 217.4; DB 8; Length 1560;  
Best Local Similarity 59.4%; Pred. No. 1.3e-25;  
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;  
Db 62 GTCCAAAAGTATTTGACATATGAGGGGTGTAAGCCTATATCGACCATCAT 121  
368 GGCCTAAGACATATTTCTGTGACTATATGCTGTGTATAGGCACTTACAGACCTTCA 427  
Qy 122 TATTAGCAACATTTAAGAACCCAGTATGATCGACCGTATTAATGTACAGTGACG 181  
428 TTTTGACTGAACCAATTAAGCGTACATCAGGGTTA--AGAGCACTTCAGTGTATA 484  
Qy 182 ATTGTGATTAAGCATTTTTCAGAAAATCATATTGGAACAATTTGTATCATATTCG 241  
485 AGGTGTGAAAATCTTGTTAAAGAGTCACTTAAGAGACACTTGTATATCGATTCG 544  
Qy 242 AAAAAAACCATTGCTTCACTGTGTGTGTAAGGGGTTAATCTCGAACACATCTGA 301  
545 ATAGCAAAACATTCATATTTCTTATTTGTAAGAAAGAGTGAAGCATCGCCAGCACTGA 604  
Qy 302 AAAGACATGAATCAACCATATCAAAAGTCAATTTAATGTACATTTGAAATTTGTCAAG 361  
605 AGGACACGAGTATGACCATATCAAAATCTTCAATTTGTCCAGAGAGAGATGCAACTCC 664  
Qy 362 CATTTTAAACATCAATCTTTAAGACATCATATTTATCTGTTCAAGAAAAACATTA 421  
665 GATTCCTAAGCATCCCAATTTAAGGGCAATTTTATCTGTTCAATTTACATMACTAA 724  
Qy 422 CGTGAACATGATTAATAGTTTTCATCGACCTTCAAAATTTAGCAACATTAATTA 481  
725 CTTGTCAACATGCAATTAAGCTTTTCAAGAGCCATATAGGCTTAAGAAATCACTTTCTA 784  
Qy 482 AACATCA-----TGTGTATCTCTGCTTATCAATGTGATCATCTGTTGTTTAA 535  
785 AACATCAAGATCCGAGGTGAGAAATCTTACCAATGTACTTTGCTGTTGTTGCAAG 844  
Qy 536 ATTTCAAACTGCTGATTTTACATTTCAATTAACAACATGATCCAAAACCTTAAT 595  
845 AGTTTCGATATGCTCACAATGTGCAATGCAATTAATAAATGATCATCTAAATTAAT 904  
Qy 596 GTCTTAATATGTGTAAGTGTGTGTAAGAAAAAGTTTATCTTCAATATGTAAAGTC 655  
905 GTCTTAATTTGTAGAAAACATGTGTGAGGAAAAATGTTTACAAATGCAATGATTTTTC 964  
Qy 656 ATGATGATTTCAACATGATCAAAATATGATGATTTGATGATGTGAGGAAATTTG 715  
965 ATGACGATCACTGATGATCAAAATATGATGATGATGATGATGATGATGATGATGAT 1024  
Qy 716 CAAAGAAAATGAATTAAGTGAACATTAAT 744  
Db 1025 CTAGAAAAACATGATCTTCTCAGCATTAAT 1053

RESULT 10  
YSCFIIIA 1739 bp DNA linear PLN 13-SEP-1996  
LOCUS YSCFIIIA  
DEFINITION *Saccharomyces cerevisiae* transcription factor IIAA (TFIIIA) gene,  
complete cds.  
ACCESSION M80611

VERSION M80611.1 GI:172902  
KEYWORDS transcription factor IIAA, zinc-finger protein, zinc-finger  
transcription factor  
SOURCE *Saccharomyces cerevisiae* (baker's yeast)  
ORGANISM *Saccharomyces cerevisiae*  
Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;  
*Saccharomycetales*; *Saccharomycetaceae*; *Saccharomyces*.  
REFERENCE  
1 (bases 1 to 1739)  
Archambault, J., Milne, C.A., Schappert, K.T., Baum, B., Friesen, J.D.  
and Segall, J.E.  
The deduced sequence of the transcription factor TFIIIA from  
*Saccharomyces cerevisiae* reveals extensive divergence from *Xenopus*  
TFIIIA  
J. Biol. Chem. 267 (5), 3282-3288 (1992)  
JOURNAL J. Biol. Chem. 267 (5), 3282-3288 (1992)  
MEDLINE 92147684  
PUBMED 1737784  
FEATURES  
source  
1. .1739  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="W303-1A"  
/db\_xref="taxon:4932"  
/chromosome="XVI"  
/gene="TFIIIA"  
/protein\_id="P12903"  
/product="transcription factor IIAA"  
/translation="MGKLVNNEGMLPRLKQRTPIRSSESSLSNLSLSTRSSN  
RPTVFCYDCDKATPRLSILRHEVTHQSLRARPQCDKASFKKSHLEHLYTH  
SDTKPRQCYGCGVTRPQRLKHEVTHQSLRARPQCDKASFKKSHLEHLYTH  
HKLTCCHNKSFORPYRLRNHLSKHDPVENYVCTTFAAGCCKEFRISQLOSHIND  
HPLKCPISKSPGVENGLOMHMIDHDSLVTKMKCHLCPDMSFKRKHDLTHYGI  
HREEDIPLIKYKISDIOQLVDHGVGLNSKRSNODDEKXISNLRKRRKXLENNV  
EFLQNEVDLEKRLSESGENGLNLNTVGKRYKSPFNRCSTRTFTKKEIKRHDKHNV  
ELKILQREKENTLVQNHKEPFIIOKETQAGDK"

ORIGIN  
Query Match 17.6%; Score 217.4; DB 8; Length 1739;  
Best Local Similarity 59.4%; Pred. No. 1.3e-25;  
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;  
Db 62 GTCCAAAAGTATTTGACATATGAGGGGTGTAAGCCTATATCGACCATCAT 121  
367 GGCCTAAGACATATTTCTGTGACTATATGCTGTGTATAGGCACTTACAGACCTTCA 426  
Qy 122 TATTAGCAACATTTAAGAACCCAGTATGATCGACCGTATTAATGTACAGTGACG 181  
427 TTTTGACTGAACCAATTAAGCGTATCAAGGGTTA--AGAGCACTTCAGTGTATA 483  
Qy 182 ATTGTGATTAAGCATTTTTCAGAAAATCATATTGGAACAACATTTGTATCATATTCG 241  
484 AGGTGTGAAAATCTTGTTAAAGAGTCACTTAAGAAACCTGTTATATGATGATTTTC 543  
Qy 242 AAAAAAACCATTGCTTCACTGTGTGTGTAAGGGGTTAATCTCGAACACATCTGA 301  
544 ATAGCAAAACATTCATATTTCTTATTTGTAAGAAAGAGTGAAGCATCGCCAGCACTGA 603  
Qy 604 AGGACACGAGTATGACATATCAAAATCTTCAATTTGTCCAGAGAGATGCAACTCC 663  
362 CATTTTAAACATCAATCTTTAAGACATCATATTTATCTGTTCAAGAAAACATTA 421  
664 GATTCCTAAGCATATTCATTAAGGGCAATATTTATCTGTTCAATTTACATTAATCTTA 723  
Qy 422 CGTGAACATATTAATTAAGTTTTCATCTGACCTTCAAAATTTAGCAACAATTAATTA 481  
724 CTTGTCAACATGCAATTAAGCTTTCAAGGGCAATATAGGCTTAAGAAATCACTTTCTA 783

Oy		482	AACATCA-----TGGTGCATCTCGTGTTGTCATGTGATCATCTCGTGTGTTAA	535
Db		784	AACATCAGATCTCGAGGTGAAGAAATCTTACCATTACTTTGCTGTTGCCAAG	843
Oy		536	ATTCCAAACTTGGTCACTATTACAAITTCATATAAACCAACTGATCCAAAATTAAT	595
Db		844	AGTTCCGATATGGTCACAATTGCATATGCATATATAAAGAATCATCTTAAATTAAT	903
Oy		596	GTCCTAATATNGTTAAAGTGTGTGTTGGAAAAAGTTATCTTCATATNGTTAATC	655
Db		904	GTCCTATTGTAGCAACCACTGTGTGGGGAAAAAGTTTCAAAATGCAATGATTATTC	963
Oy		656	ATGATGATTTACACATGATCAAAATATGACCTTGTGATTATGTGATGTGGGAAATTTG	715
Db		964	ATGACGATCACTACTGTAACCAAAAATTGGAAGTGTGATATATGTCATATGTCCTTTTT	1023
Oy		716	CAAGAAAAATGATTAAGTTACTGAACATTAT	744
Db		1024	CTAGAAAAACATGATCTTCTCACGCATTAT	1052

RESULT 11				
YSCP9677/c				
LOCUS	YSCP9677	37497 bp	DNA	linear
DEFINITION	Saccharomyces cerevisiae chromosome XVI cosmid 9677.	PLN 01-AUG-1997		
ACCESSION	U25841 U00094			
VERSION	U25841.1 GI:786295			
KEYWORDS				
SOURCE				
ORGANISM	Saccharomyces cerevisiae (baker's yeast)			
	Saccharomyces cerevisiae			
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
REFERENCE				
AUTHORS	1 (bases 1 to 37497) Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Du,Z., Favello,A., Fulton,L., Gattung,S., Greco,T., Kirsten,J., Kucaba,T., Hallsworth,K., Hawkins,J., Hillier,L., Jier,M., Johnson,D., Johnston,L., Langston,Y., Latreille,P., Le,T., Mardis,E., Menezes,S., Miller,N., Nhan,M., Pauley,A., Peluso,D., Rifken,L., Riles,L., Tatch,A., Trevaslis,E., Vignati,D., Wilcox,L., Wohlman,P., Vaudin,M., Wilson,R. and Waterston,R. The sequence of a portion of the right arm of Saccharomyces cerevisiae chromosome XVI Unpublished (1995) Miller,N. 2 (bases 1 to 37497) The sequence of S. cerevisiae cosmid 9677 Unpublished (1995) 3 (bases 1 to 37497) Waterston,R. Direct Submission Submitted (27-Apr-1995) Robert Waterston Submitted by: Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA e-mail: mj@sequencer.wustl.edu			
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

NEIGHBORING COSMID INFORMATION:

This sequence includes nucleotides 1-37497 of cosmid 9677. The  
cosmid on the left is 9705.

Location/Qualifiers

1..37497

/organism="Saccharomyces cerevisiae"

/mol\_type="genomic DNA"

/strain="S288C (AB972)"

/db\_xref="taxon:4932"

/map="XVI"

/note="baker's yeast"

complement(2366)..3655

/gene="YFC2"

complement(2366)..3655

/gene="YFC2"

CDS

[illegible]



AUTHORS Archambault, J., Schappert, K.T. and Priesen, J.D.  
 TITLE A suppressor of an RNA polymerase II mutation of *Saccharomyces cerevisiae* encodes a subunit common to RNA polymerases I, II, and III  
 JOURNAL Mol. Cell. Biol. 10 (12), 6123-6131 (1990)  
 MEDLINE 91061718  
 PUBMED 2247052  
 COMMENT Original source text: *S.cerevisiae* single-copy plasmid DNA, clone 2.  
 Draft entry and computer-readable sequence for [unpublished (1990)] kindly submitted by J.D.Priesen, 01-MAY-1990  
 The Hospital for Sick Children  
 555 University Avenue  
 Toronto, Ontario, CANADA M5G 1X8  
 e-mail: jim@ckids.toronto.edu.  
 Location/Qualifiers  
 1..2133  
 /organism="Saccharomyces cerevisiae"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4932"  
 join(1026..1045,1122..1569)  
 /note="RNA polymerase II sixth subunit (RP026)"  
 /codon\_start=1  
 /protein\_id="AAA34989.1"  
 /db\_xref="GI:172453"  
 /translation="MSDYEAFAFDNGENFEDVEHFSDEETYEKPKQGETTDAN  
 GRTIVGNGPDPFOHBOIRRTLEKKAIPDQATPTPYMKREARLIGRAQIS  
 MNAPVFDLEGEDPLRIAMKLEKIPVIRRLPDSFEDMSVELIYDL"  
 <1026..1045  
 /note="RNA polymerase II sixth subunit (RP026)"  
 /number=1  
 1046..1121  
 /note="RP026 intron A"  
 1122..>1569  
 /note="RNA polymerase II sixth subunit (RP026)"  
 /number=2  
 ORIGIN Chromosome XVI.  
 Query Match 16.1%; Score 199.2; DB 8; Length 2133;  
 Best Local Similarity 61.4%; Pred. No. 1.1e-22;  
 Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;  
 QY 171 TACAGTGCAGATTGTGATTAAGCATTTTTCAGAAAATACATTTGGAAACACATATTGT 230  
 Db 556 TTCAAGTGTATAGTGCAAAATCCTTCGTTAAAAGAGTCACTTGAAGACACTTGA 497  
 QY 231 ATCACATTCCGAAAAAAACCATTCATTGTTCAAGTGTGTGTAAGGGGTTAATTCCTG 290  
 Db 496 TACGCAATTCGTATACGAAACCAATGTTCTTATGTGAAAAAGAGTACGACTCG 437  
 QY 291 ACAACCTGTAAGAGCATGAATACCCGATCAAAAGCATTTAATTAATGTAACATTTGAAA 350  
 Db 436 CCAAGCACTGAAGCGACAGAGTAAGCATCAAAATTTTCATTTGTCAGAGAAAG 377  
 QY 351 TTGTCAAGAGCATTTTATTAACATCAATCTTTAAGACATCAATATTAATTCGTTGATGA 410  
 Db 376 ATGCAACCTCGATTTTACAGATCCACATTTAAGGCAATATTTATCTGTTCAATT 317  
 QY 411 AAAAAATTAAAGTGTAAACATGTAATAAGTTTCACTGACCTTCAAAAATTAGACA 470  
 Db 316 ACATAACTAAACCTGTCACACATGCAATAAAGCTTTCAGAGGCCATATAGCTAAGAAA 257  
 QY 471 ACATAAATTAAACATCA-----TGTGTGATCTCTGCTTATCAATGTGATCATCTCG 524  
 Db 256 TCACTATTCTAAACATCAAGATCTGAGTAGAAAAATCTTACCAATGTATCTTTGCTCG 197  
 QY 525 TTGTTTAAAAAATTTCCAACTGTGTCAGTATTAACAATTTCAATTAACAACATGCAATCC 584  
 Db 196 TTGTTGCAAGAAGTTTCGATATGTCACAAATTGCAATGGCATATAAAAAAGATCATCC 137  
 QY 585 AAAACTTAATGTCTTAATGTGTAAAGGTTGTGTGGAAAAAAGTTTATCTTCACA 644

Db 136 TAAATTAAGTCTTATTTAGCAAAACATGTGTGGGGAAAAATGGTTACAATGCA 77  
 QY 645 TATGTTAAGTCATGATGATTTTACCATGATCAAAATATGACTTGTGATTAATGTGATGT 704  
 Db 76 CATGATTTATTCATGACGACTACTAGTAGTAACCAAAATTTGAAGTGTCAATATATGTCCTGA 17  
 QY 705 GGGGAATTTGCA 718  
 Db 16 TATGCTTTTCTA 3  
 RESULT 13  
 CDS061NG  
 LOCUS  
 DEFINITION T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533  
 ACCESSION AL400514  
 VERSION AL400514.1 GI:12156635  
 KEYWORDS  
 ORGANISM  
 Saccharomyces bayanus  
 Saccharomyces bayanus  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetaceae; Saccharomycetes.  
 1 (bases 1 to 797)  
 Boulet, J.L., Algle, M., Artiguenave, F., Blandin, G.,  
 Boulet, J.L., Algle, M., Bon, E., Brotier, P., Casaregola, S.,  
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
 Malpercuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Poirier, S.,  
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
 Wincker, P. and Weissenbach, J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 FEBS Lett. 487 (1), 3-12 (2000)  
 11152876  
 2 (bases 1 to 797)  
 Bon, E., Neuvéglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,  
 Algle, M. and Durrens, P.  
 Genomic exploration of the hemiascomycetous yeasts: 5.  
 Saccharomyces bayanus var. uvarum  
 FEBS Lett. 487 (1), 37-41 (2000)  
 20584715  
 11152880  
 3 (bases 1 to 797)  
 Genoscope.  
 Direct Submission  
 Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,  
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.  
 Location/Qualifiers  
 1..797  
 /organism="Saccharomyces bayanus"  
 /mol\_type="genomic DNA"  
 /strain="CLIB 533"  
 /variety="uvarum"  
 /db\_xref="taxon:4931"  
 /clone="AS0AA022B07"  
 /clone\_1ib="AS0AA"  
 /note="end : 77"  
 <2..>796  
 /note="similar to Saccharomyces cerevisiae ORF YPR186c [  
 PZF1 / TFI1A (transcription initiation factor) ]"  
 /evidence="not\_experimental"



ORIGIN

Query Match 15.6%; Score 193; DB 11; Length 797;

Best Local Similarity 59.3%; Pred. No. 1.3e-21;

Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;

QY 178 GACGATGTGATTAAGCAATTTTCAGAAAATCAATTGGAAAACATATTTGATCAT 237  
 DB 83 GACAAATTGTCAAAATCTTTGTTAAABAAGATCATTAAGAGGACCTGTTTGGCAT 142  
 QY 238 TCCGAAAAAAACCAATTCATTTGTCAGTGTGTGTAAAGGGTTAATTTCCACAAC 297  
 DB 143 TCTGATACGAAGGCATTCATGTTCTTATGTGGCAAGGGGTGAACAATCCACAGCA 202  
 QY 298 TTGAAAGACATGAATTCACCAATACAAAGCTTTAAATGATCATTTGAAATTTGCA 357  
 DB 203 TTGAAGGCCACGAAGTTACACATCCAAATCTTTGTTGTCCGAAAGATGATGAT 262  
 QY 358 GAAGCATTTTAAATCAATCATCTTTAAAGACATCATATTTATCTGTTCAAGAAAAACA 417  
 DB 263 CTCGATTTTAAAGCACCCTCAATTAAGGGCTCATTTCTATCTGTCCACCTAGACAA 322  
 QY 418 TTAACGTGAACAATGAATTAAGTTTCACTGACCTTCAAAATTAAGCAACAATAA 477  
 DB 323 TTAACCTGTCCAAATGTAGTAAATTTTCAAGAGACCTTACAGGCTAAGAAATCATATC 382  
 QY 478 TTTAAACATCATGTG-----GGATCTCTGCTTATCAATGATCATCTGCTGTTT 531  
 DB 383 TCTAAACATCATATCTGTAAGGTGTAATCCGATCAATGATCTTCCGCTGCTGTGT 442  
 QY 532 AAAAATTTCCAACTTGAGTATTAATTTCAATTAATAACAATGATCCAAAATCT 591  
 DB 443 ACAGAGTACCGTATATGTATGATGATGATGATGATGATGATGATGATGATGATG 502  
 QY 592 AAATGCTTAATATGTGTAAGGTGTGTGGGAAAAAGTTTATCTTCAATATGTA 651  
 DB 503 AAATGCCCCATTGCAACAAGCCTTGTGTAGGGGAAATGATCTTCAAAATGATATGATC 562  
 QY 652 AGTATGATGATTTTACATGATGATGATGATGATGATGATGATGATGATGATG 711  
 DB 563 ATTATGATGATGATTTTGTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 622  
 QY 712 TTTGCAAGAAAAATGATATGATGATGATGATGATGATGATGATGATGATG 756  
 DB 623 TTTCTGGAAGAACGACCTCTTACTACTACTACTACTACTACTACTACTACTACT 667

RESULT 14  
 AE016818\_01

WPCOMMENT

Sequence split into 16 fragments LOCUS AE016818 Accession AE016818

Fragment Name	Begin	End
AE016818_00	1	110000
AE016818_01	100001	210000
AE016818_02	200001	310000
AE016818_03	300001	410000
AE016818_04	400001	510000
AE016818_05	500001	610000
AE016818_06	600001	710000
AE016818_07	700001	810000
AE016818_08	800001	910000
AE016818_09	900001	1010000
AE016818_10	1000001	1110000
AE016818_11	1100001	1210000
AE016818_12	1200001	1310000
AE016818_13	1300001	1410000
AE016818_14	1400001	1510000
AE016818_15	1500001	1519138

Continuation (2 of 16) of AE016818 from Base 100001 (AE016818 Eremothecium gossypii ATCC  
 Query Match 9.8%; Score 120.6; DB 8; Length 110000;  
 Best Local Similarity 50.1%; Pred. No. 1.9e-10;  
 Matches 386; Conservative 0; Mismatches 374; Indels 11; Gaps 3;

QY 25 TCGATATCATCTTTAATATCTTCTTCTCATCATGCTCCCAAAAGTATATTTGCACA 84  
 DB 17349 TCGTTCATGAGTGTATGATTCACAGCGGTGAGACACCGGCGCAAAACGATTTCTGCGAC 17408  
 QY 85 TATGAAGGTGTGATTAAGCCTTAATATGACATCATTTATTAAGCAACA-TTTAAGAAC 143  
 DB 17409 TACGAGGCTGCTTACAAAGCGTTCACGCGCGGTCCCTGCTGACGAGACACGACAGACG 17468  
 QY 144 CCACAGTATGATCGACCGTATTAATATGATGATGATGATGATGATGATGATGATG 203  
 DB 17469 GCGCACACAGGGAATCGAGGCAATACAGTGC---GAGCATGTGGGCGGGTTTACAA 17524  
 QY 204 AAAATCACATTTGGAAACAATATGATATCACTTCCGAAAAAAACATTCATTTGTC 263  
 DB 17525 GAGTGTGCACTTGGACCGGCACTGTTCTGCACTTGTGACAAACACCTTACGCTGCAC 17584  
 QY 264 AGTGTGTGTAAGGGGTAAATTTCTGACAAACATTTGAAAGAATGAAATCAACCATAC 323  
 DB 17585 GGTATGTGCAAGGGGTGACACCGCGCAGCAATGAGGGCGGACGATGACACACAC 17644  
 QY 324 AAGTATTTTAATGATGATTTGAAATTTGTCAAGAGCATTTTATTAACATCATCTTT 383  
 DB 17645 AAGTGTTCAGTGTCCGACGAGGGCTGCGCGGAAGGTTCTTACAAAGCACCCGAGTT 17704  
 QY 384 AAGACATCATATTTATCTGTTATGATGAAAAAATTAAGTGAACATGTAATAAGT 443  
 DB 17705 GCGTTCGACGCTGCTGCGGTGACGACGAAAGCTGATGATGATGATGATGATG 17764  
 QY 444 TTTCACTGACCTTCAAAATTAAGCAACAATTAATTAATTAATTAATTAATTAATTA 500  
 DB 17765 GTTCCAAAGGCAATACGTTTGAAGAACACATCGCAAGCATCAAGGGGCTGATCGCA 17824  
 QY 501 ---TGCTTATCAATGATGATCATCTGTTGTTTAAATTTGCAAACTTGTGATAT 557  
 DB 17825 GTTCCGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 17884  
 QY 558 ACAATTTCAATTAATTAACAATGATGATGATGATGATGATGATGATGATGATGATG 617  
 DB 17885 GCGACGACCTTCAACGAGCACCGGAAGCTACATGATGATGATGATGATGATGATG 17944  
 QY 618 TGTGGGAAAAAAGTTTATCTTCAATATGATGATGATGATGATGATGATGATGATG 677  
 DB 17945 CGTGGGGGAGACGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 18004  
 QY 678 AATATGACTTGTGATTTATTTGATGATGATGATGATGATGATGATGATGATGATG 737  
 DB 18005 AAATGGAAGTGTCTGTGTGTGATGATGATGATGATGATGATGATGATGATGATG 18064  
 QY 738 ACATTTAATATCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 788  
 DB 18065 GCATTTATGAAAAACCGATGATGATGATGATGATGATGATGATGATGATGATG 18115

RESULT 15

AC138126/c

LOCUS AC138126 169063 bp DNA linear PRI 17-DEC-2002

DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.

AC138126

VERSION AC138126.1 GI:2751357

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 169063)

DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 169063)

DOE Joint Genome Institute and Stanford Human Genome Center.

AUTHORS Direct Submission

JOURNAL Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell



## COMMENT

Drive, Walnut Creek, CA 94598, USA  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.

## FEATURES

## source

1. 169063  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="RP11-274A19"

## ORIGIN

Query Match 8.2%; Score 101; DB 9; Length 169063;

Best Local Similarity 50.8%; Pred.No.2.4e-07; Mismatches 280; Indels 9; Gaps 2;

Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATGATATCATCTTTAATATCTTCTTCATACAGTCCC 66  
|||  
Db 80134 GAATGGGCAAGCTTTTAAACAGTCTCATCTTACTACACATAGAGAAATCATACT 80075  
|||  
QY 67 AAAAAATATTTGCAATATGAGGGTGTGATAAAGCTATATGACCATCATATTATTA 126  
|||  
Db 80074 GGAGGAAATCCTACAAATGTGAGAAATGGCAAGCTTCTATGATCTCAAAACTT 80015  
|||  
QY 127 GAGCAACATTTAAGAACCCAGTATATGACCGTATAAATGTACAGTGAAGATTGT 186  
|||  
Db 80014 ACTGAACATAGAAATTTATCTAGAGAAACCTTACACAGT-----GAGAGATGT 79961  
|||  
QY 187 GATTAAGCATTTTTCAGAAATACATTTGAAACATATTTGTATCATTCGGAATA 246  
|||  
Db 79960 GGCAGAGCTTTTACCATCTTCCACATCTGTCTACACATAGATATTCATCTGAGAG 79901  
|||  
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCTGACACACTTGAAGA 306  
|||  
Db 79900 AAACCTTACCAATGTGAGAAATGTGTAAAGCTTTAACAGTCTTCAACCTTACTGA 79841  
|||  
QY 307 CATGAATGACCATATCAAAAGTCATTTAATGTACATTTGAAATTTGTCAAGAGCATTT 366  
|||  
Db 79840 CATTAAGAAATTCATATCTGAGAGAAACCTTACCAATGTGAAAAATGTGCAAGCTTTT 79781  
|||  
QY 367 TATTAACATCAATCTTTAA---GACATCATATTTATCTGTCAAGAAATAACATTAACG 423  
|||  
Db 79780 AACAGTCTCTCAACCTTACTGAGACATAAGAAATTCATCTGTGTAGAAACTTACAAA 79721  
|||  
QY 424 TGTAAACAATGTATTAAGTTTCACTGACCTTCAAAATTAGCACACATTAATTAAAA 483  
|||  
Db 79720 CCTAAAGATGTGACAGTGTATTTGAAAAACCTTCAAAGTTTCTTAAACATAAAGAAAT 79661  
|||  
QY 484 CATCATGTGATCTCTCTGCTATCATGTGATCATCTCGTGTGTTTAAATTTCCAA 543  
|||  
Db 79660 TATGCTGGTGAAGAAATCTTGAAGATGTGAAGATGTACAAAACCTTAAAGTGTCTAC 79601  
|||  
QY 544 ACTTGGTCAGATTACATTTTCATATTAACAACATGATCCAAAATCT 590  
|||  
Db 79600 ACTTGAATGTGATAGATATTCATTCATTAATAAAAAAACTTCAAGT 79554  
|||

Search completed: February 9, 2005, 05:36:20  
Job time : 5288.09 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:43:04 ; Search time 640.722 Seconds  
(without alignments)  
11419.610 Million cell updates/sec

Title: US-09-831-804-1\_COPY\_720\_1955

Perfect score: 1236  
Sequence: 1 atgagtgagagacgacgaac.....aaacatcagtgattctcga 1236

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1236	100.0	2060	3	AAA15398	AAA15398 DNA encod
2	1232.8	99.7	1239	6	ABZ32206	ABZ32206 Candida a
3	99.4	8.0	2132	10	ADA53158	ADA53158 Human cod
4	96.8	7.8	1705	9	ACA98970	ACA98970 cDNA enco
5	96.6	7.8	2320	10	ACA56456	ACA56456 Human sig
6	96.6	7.8	2320	12	AD156252	AD156252 Human pol
7	95.2	7.7	4563	5	AA591317	AA591317 DNA encod
8	95	7.7	981	12	ADN04717	ADN04717 Antipsoi
9	94.6	7.7	2597	8	AA151569	AA151569 Human nuc
10	92.4	7.5	3639	5	AA564586	AA564586 DNA encod
11	92	7.4	994	12	ACH87568	ACH87568 Human gen
12	92	7.4	2873	13	ADR14368	ADR14368 Human NF-
13	91.8	7.4	1533	12	ADN99064	ADN99064 Novel hum
14	91.8	7.4	1533	12	ADN00633	ADN00633 Novel hum
15	91.8	7.4	2110	10	ADDB3624	ADDB3624 Human cDN
16	91.8	7.4	2760	10	ADCS6695	ADCS6695 Human mac
17	91.8	7.4	3078	10	ADBE2468	ADBE2468 Human cDN
18	91	7.4	3755	13	ADR07315	ADR07315 Full leng
19	91	7.4	4209	13	ADR08223	ADR08223 Full leng
20	90.4	7.3	2026	8	AA55863	AA55863 Human nuc

21	90.4	7.3	2114	10	ADA53124	ADA53124 Human cod
22	90.2	7.3	2230	10	ADA52931	ADA52931 Human cod
23	89.8	7.3	590	12	ACH80103	ACH80103 Human gen
24	89.2	7.2	2064	10	ADC30762	ADC30762 Human nov
25	89.2	7.2	2221	11	ADM01344	ADM01344 Human cDN
26	89.2	7.2	2632	5	AA568872	AA568872 DNA encod
27	89.2	7.2	2729	4	AAH16178	AAH16178 Human cDN
28	89.2	7.2	3839	6	ABK83826	ABK83826 Human cDN
29	89.2	7.2	3839	12	ADN04207	ADN04207 Antipsoi
30	89.2	7.2	3839	13	ADR25049	ADR25049 Breast ca
31	89	7.2	1991	13	ADN09824	ADN09824 Human the
32	89	7.2	2298	4	AAH16608	AAH16608 Human cDN
33	89	7.2	2299	13	ADN09825	ADN09825 Human the
34	89	7.2	2905	5	AA566143	AA566143 DNA encod
35	89	7.2	3003	12	ADN06033	ADN06033 Antipsoi
36	89	7.2	3020	5	AA592560	AA592560 DNA encod
37	89	7.2	3502	4	AA157845	AA157845 Human pol
38	88.6	7.2	1335	12	ADN98906	ADN98906 Novel hum
39	88.6	7.2	1335	12	ADN00475	ADN00475 Novel hum
40	88.6	7.2	1466	12	ACH91879	ACH91879 Human gen
41	88.6	7.2	1875	10	AD160591	AD160591 Secreted
42	88.6	7.2	2597	8	AA558855	AA558855 Human nuc
43	88.6	7.2	4227	9	ACA98938	ACA98938 cDNA enco
44	88.4	7.2	1410	12	ACH87337	ACH87337 Human gen
45	88.4	7.2	1952	10	ADC58104	ADC58104 Zinc fing

## ALIGNMENTS

RESULT 1	
AAA15398	
ID	AAA15398 standard; DNA; 2060 BP.
XX	
AC	AAA15398;
XX	
DT	04-SEP-2000 (first entry)
XX	
DE	DNA encoding a transcription factor designated CATFIIIA.
XX	
KW	Transcription factor; CATFIIIA; DNA-binding protein;
KW	ribosomal RNA 5S gene; fungal infection; ss.
XX	
OS	Candida albicans.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	720..1958
FT	/*tag= a
FT	/transl_except= (pos: 1296..1298, aa: Ser)
FT	/transl_except= (pos: 1734..1736, aa: Ser)
XX	
PN	WO200028037-A1.
XX	
PD	18-MAY-2000.
XX	
PF	09-NOV-1999; 99WO-FR002739.
XX	
PR	10-NOV-1998; 98FR-00014147.
XX	
PA	(HMRI ) HOECHST MARION ROUSSEL.
XX	
PI	Bordon-Pallier F, Camier S, Sentenac A;
XX	
XX	WPI; 2000-376549/32.
DR	P-PSDB; AAY93316.
XX	
PT	New nucleic acid encoding Candida albicans transcription factor, useful
XX	e.g. in screening for antileukocytic agents and for immunization.
XX	Claim 4; Page 32-33; 45pp; French.
PS	
XX	The present sequence encodes a Candida albicans transcription factor,
CC	designated CATFIIIA. The polypeptide is a DNA-binding protein, which is

CC involved in initiating transcription of the ribosomal RNA 5S gene. The  
CC polynucleotide is used to screen for its specific inhibitors, potentially  
CC useful as antimycotic agents, to raise an antibody response that is  
CC protective against fungal infection and to raise antibodies. Such  
CC antibodies, as well as the polypeptides and polynucleotides are used in  
CC compositions for diagnosing and treating fungal infections, e.g. by  
CC detecting polymorphisms and mutations

XX Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 1236; DB 3; Length 2060;

Best Local Similarity 100.0%; Pred. No. 1.4e-233; Mismatches 0; Indels 0; Gaps 0;

Matches 1236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTGAAGTGAAGCAAAACCAATGCATATCATCTTAAATCTTCTTCTTCATCA 60  
DB 720 ATGAGTGAAGTGAAGCAAAACCAATGCATATCATCTTAAATCTTCTTCTTCATCA 779  
QY 61 CGTCCCAAAAAGTATTTTGGACATATGAAAGGTGATTAAGCCATTAATGCACATCA 120  
DB 780 CGTCCCAAAAAGTATTTTGGACATATGAAAGGTGATTAAGCCATTAATGCACATCA 839  
QY 121 TTATTAGACAACATTTAAGAACCCACAGTAATGATCGACCGTAAATAGTACAGTGAC 180  
DB 840 TTATTAGACAACATTTAAGAACCCACAGTAATGATCGACCGTAAATAGTACAGTGAC 899  
QY 181 GATTGTGATTAAGCATTTTTCAGAAATGACATTTTGGAAAACATATTGATCAATCC 240  
DB 900 GATTGTGATTAAGCATTTTTCAGAAATGACATTTTGGAAAACATATTGATCAATCC 959  
QY 241 GAAAAAAAACATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACCTTG 300  
DB 960 GAAAAAAAACATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACCTTG 1019  
QY 301 AAAAGACATGAATCAACCATTAAGTCAATTTAAATGTACATTTGTAATGTCAAGAA 360  
DB 1020 AAAAGACATGAATCAACCATTAAGTCAATTTAAATGTACATTTGTAATGTCAAGAA 1079  
QY 361 GCATTTTAAACATCATCTTAAAGACATCATATATTAATCTGTCATGAAACATTA 420  
DB 1080 GCATTTTAAACATCATCTTAAAGACATCATATATTAATCTGTCATGAAACATTA 1139  
QY 421 AGGTGTAACATGTAAATGAAGTTTCACTCGACCTTCAAAATTAGACACATTAATTA 480  
DB 1140 AGGTGTAACATGTAAATGAAGTTTCACTCGACCTTCAAAATTAGACACATTAATTA 1199  
QY 481 AAAACATCATGTGATCTCTGCTTATCATGTGATCATCTGCTTATTAATTTTC 540  
DB 1200 AAAACATCATGTGATCTCTGCTTATCATGTGATCATCTGCTTATTAATTTTC 1259  
QY 541 CAAACTTGGTCAATTTACATTTTCAATTAACAATCAGTCACCAAACTTAATGCTCT 600  
DB 1260 CAAACTTGGTCAATTTTACATTTTCAATTAACAATCAGTCACCAAACTTAATGCTCT 1319  
QY 601 AAATGTGTAAAGTTGTGTGGAAAAAGTTTATCTTCAATATGTTAACTCATGAT 660  
DB 1320 AAATGTGTAAAGTTGTGTGGAAAAAGTTTATCTTCAATATGTTAACTCATGAT 1379  
QY 661 GATTTCACATGATCAAAATATGAGACTTGATTAATGATGATGGGAAATTTGCAAG 720  
DB 1380 GATTTCACATGATCAAAATATGAGACTTGATTAATGATGATGGGAAATTTGCAAG 1439  
QY 721 AAAAATGAATTAAGTGAACATTTATTAATCTTCATGATGATGTAATCCCTGATGATTA 780  
DB 1440 AAAAATGAATTAAGTGAACATTTATTAATCTTCATGATGATGTAATCCCTGATGATTA 1499  
QY 781 TTAAGGAAATCTGAAGTGAAGAAATTAAGAACTTAATGATCAAGATCGAAATTAAT 840  
DB 1500 TTAAGGAAATCTGAAGTGAAGAAATTAAGAACTTAATGATCAAGATCGAAATTAAT 1559  
QY 841 AATTTCGATGATTAAGAAACAGAAATTAAGTGAAGAGATGAAGAAATGAAGAA 900  
DB 1560 AATTTCGATGATTAAGAAACAGAAATTAAGTGAAGAGATGAAGAAATGAAGAA 1619

QY 901 GATAGCTGATGATGAAAAAAGATGATGTAGATCAGACTCATGTCAAGTCAAGATCA 960  
DB 1620 GATAGCTGATGATGAAAAAAGATGATGTAGATCAGACTCATGTCAAGTCAAGATCA 1679  
QY 961 ATTAATCATCTTACTGCTTCTTTGGAAGTTCAAAGATGTTTCTTAATCTTATGAT 1020  
DB 1680 ATTAATCATCTTACTGCTTCTTTGGAAGTTCAAAGATGTTTCTTAATCTTATGAT 1739  
QY 1021 AGTGGAAAGATCAATTTGCTCTAAGATAATTTGTATAGATGTTTCTAAGAAATAT 1080  
DB 1740 AGTGGAAAGATCAATTTGCTCTAAGATAATTTGTATAGATGTTTCTAAGAAATAT 1799  
QY 1081 GATTACGTCGACATTTGAAATGGCATGATGATTAATTAACAAAGATGATGATCTTA 1140  
DB 1800 GATTACGTCGACATTTGAAATGGCATGATGATTAATTAACAAAGATGATGATCTTA 1859  
QY 1141 AATAGTATGAAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1200  
DB 1860 AATAGTATGAAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1919  
QY 1201 GATTATTGCCAAATGAATCAATCAATGATTTCTCGA 1236  
DB 1920 GATTATTGCCAAATGAATCAATCAATGATTTCTCGA 1955

## RESULT 2

ID AB232206  
AB232206 strand; DNA, 1239 BP.

XX AB232206;

XX 30-JAN-2003 (first entry)

DE Candida albicans essential gene SEQ ID NO 6493.

KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;

KW signal transduction; DNA replication; cell division; growth;

KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.

OS Candida albicans.

PN WO200253728-A2.

PD 11-JUL-2002.

PP 26-DEC-2001; 2001MO-US049486.

PR 29-DEC-2000; 2000US-0259128P.

PR 20-FEB-2001; 2001US-0079202A.

PR 22-AUG-2001; 2001US-0314050P.

PA (ELIT- ) ELITRA PHARM INC.

PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,

DR WPI; 2002-566694/60.

DR P-PsDB; ABP73656.

PS Claim 37; SEQ ID NO 6493; 167bp + Sequence Listing; English.

CC The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal



PT which the gene is involved, or as target molecules for gene therapy.  
XX  
XX Claim 1; SEQ ID NO 726; 205bp; English.  
XX  
XX The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX  
SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;  
  
Query Match 8.0%; Score 99.4; DB 10; Length 2132;  
Best Local Similarity 50.6%; Pred. No. 4.7e-10;  
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;  
  
QY 7 GAAAGTGAAGAAACCAATCGATATCATCTTAAATCTCTCTCTCATCAGCTGCC 66  
DB 1384 GAATGTGGAGAGCTTTTAAACCAATCTCAATCTTACTACACATAGAGAAATTCATCT 1443  
QY 67 AAAAGTATATTTTGCATATGAGAGGTGTGATTAAGCTTATATGACCATCATATTATTA 126  
DB 1444 GGAGAGAAATCTTACAAATGTGAGAAATGTGGCAAGCTTTCTATGATCTTCAAAACT 1503  
QY 127 GAGCAACATTTAAGAAACCAAGTAATGATCGACCTATTAATGTACAGTGAAGATTGT 186  
DB 1504 ACTGAACATAGAGAAATTTATCTAGAGAGAAACCTTACACATGT-----GAGAAATGT 1557  
QY 187 GATTAAGCATTTTTCAGAAATTCATCTTGGAAACATATTTGTATCAGATTCCTCCGAAAA 246  
DB 1558 GCGAAAGCTTTTAAACCATCTTCAACCTTGTACATTAAGGTAAATTTCTACTGAGAG 1617  
QY 247 AAACATTCCTATTTTCACTGTGTGTAAAGGGTAAATTTCTGACAACTTGAAGAAA 306  
DB 1618 AAACCTTACCAATGTGAAGAAATGTGTAAAGCTTTAAACAGTCCACACCTTACTATA 1677  
QY 307 CATGAATACCCATACAAAGTCAATTAATGTACATTTGAAATTTGCAAGAGCATTT 366  
DB 1678 CATAGAGAAATTCATCTGAGAGAAACCTTACCAATGTGAAAAATGTGGCAAGCTTTT 1737  
QY 367 TATTAACATCATCTTTAA---GACATCATATATATCTGTTCATGAAAAAACATTAAG 423  
DB 1738 AACCGCTCTCAAACTTACTGTGACATTAAGAAATTCATCTGTGTAGAAACCTTACAA 1797  
QY 424 TGTAAACATGTAAATTAAGTTTCACTCCACCTTCAAAATTTAGCAACATTAATTA 483  
DB 1798 CCTAAAGATGTAAACGTGATTTTGAAGACCTTCAAAATTTCTTAAACATTAAGAAAT 1857  
QY 484 CATCATGTGTGATCTCTGCTTATCATGTATCATCTGTGTGTTTAAATTTCCAA 543  
DB 1858 TATGCTGTGTGAGAAATCTTGAAGAAATGTGAAGATGTAAACAAACCTTTAAAGTGTAC 1917  
QY 544 ACTTGTCAGTATTAACATTTTCAATTAAGAAACATGATCCAAACT 590  
DB 1918 ACTTATGTGTGATTAAGATTAATTCATTAAGAAACAAACCTTAAAGT 1964  
  
RESULT 4  
ACA98970  
ID ACA98970 strand; cDNA, 1705 BP.  
XX  
XX  
AC ACA98970;  
XX  
XX 25-JUL-2003 (first entry)  
XX  
XX cDNA encoding human nucleic acid-associated protein (NAAP) #51.  
DE Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;  
KW anti-conjunctival; nontropic; neuroprotective; cerebroprotective; anti-HIV;  
XX anti-allergic; anti-inflammatory; thyromimetic; gene therapy;  
KW cell proliferative disorder; cancer; atherosclerosis;  
KW neurological disorder; epilepsy; Huntington's disease; stroke;  
KW immune disorder; inflammatory disorder; AIDS; allergy;

KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2003023003-A2.  
XX  
XX 20-MAR-2003.  
XX  
XX  
XX 05-SEP-2002; 2002MO-US028540.  
XX  
XX  
XX 07-SEP-2001; 2001US-0317792P.  
XX  
XX 07-SEP-2001; 2001US-0317912P.  
XX  
XX 14-SEP-2001; 2001US-0322270P.  
XX  
XX 21-SEP-2001; 2001US-0324040P.  
XX  
XX 28-SEP-2001; 2001US-0326732P.  
XX  
XX 19-OCT-2001; 2001US-0346715P.  
XX  
XX 25-JAN-2002; 2002US-0351749P.  
XX  
XX 22-FEB-2002; 2002US-0359498P.  
XX  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX  
XX Tang YF, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ,  
PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BW,  
PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li  
PI Hafelia Adu, Sanjwalala B, Margulis JP, Gotrold AE, Lee SY, Ison CH;  
PI Baughn MR, Chawla NK, Nguyen DB, Swannaker A, Zebairadian Y, Shah P;  
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kabie AE;  
PI Burford N, Ramkumar J;  
XX  
XX WPI; 2003-313243/30.  
XX  
XX P-PSDB; ABU96722.  
XX  
XX  
XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
PT creating and preventing diseases or conditions associated with the  
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
XX infections.  
XX  
XX  
XX Claim 5; Page 340-341; 345pp; English.  
XX  
XX  
XX The invention describes a novel human isolated nucleic acid-associated  
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in  
CC diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression or overexpression of NAAP, such as cell  
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
CC disorders, or infections. These are also useful in assessing the effects  
CC of exogenous compounds on the expression of nucleic acid and amino acid  
CC sequences of NAAP. The NAAP or its fragments are useful in screening  
CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
CC or in altering the expression of the target polynucleotide and compounds  
CC that specifically bind to or modulate the activity of the polypeptide.  
CC The microarray is useful in monitoring or measuring protein-protein  
CC interactions, drug-target interactions, and gene expression profiles.  
CC This sequence encodes a novel human nucleic acid-associated protein  
CC (NAAP)  
XX  
XX  
SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;  
  
Query Match 7.8%; Score 96.8; DB 9; Length 1705;  
Best Local Similarity 52.9%; Pred. No. 1.5e-09;  
Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;  
  
QY 7 GAAAGTGAAGAAACCAATCGATATCATCTTAAATCTCTCTCTCATCAGCTGCC 66  
DB 695 GAATGTCCCAAAACATTTTGCATGCTTTCACACCTTAACTCAACATTAAGATCCAAACT 754  
QY 67 AAAAGTATATTTTGCATATGAGAGGTGTGATTAAGCTTATATGACCATCATATTATTA 126  
DB 755 AGAGTGAATTTCTCAACATTAAGTAAGACATTAAGAAAGCTTTAACTGTGCTCAACCTT 814

QY 127 GAGCAACATTTAAGAACCCAGTAATGATCGACCGTATTAATGTACAGTGGAGATTGT 186  
DB 815 AATTAACATTAAGAGAAATTCATCTGAGAAACCTTAACAATGTAA-----GAAATGT 868  
QY 187 GATTAAGCATTTTTCAGAAATTCATTGGAAACACATTTGTATCATCTTCGAAAA 246  
DB 869 GGCMAACCTTTTACAGACCTCACACCTTAATTAAGACATTAAGAAATTCATCTGAAAG 928  
QY 247 AAACATTCATGTTGATGTGTGTTAAAGGGGTAAATCTCGACAAACCTTGAAGA 306  
DB 929 AAACCTTAAGAAATGTAAAGATGTGGCAAGCCCTTAACACAGTATGACCTTACTACA 988  
QY 307 CATGAATCAACCCATACAAAGTCATTTAAATGTATCATTTGAAATTTGCAAGAACATTT 366  
DB 989 CATTAATTAATTAATCTGAGGAAATTCCTTAACAATGTGAGAAATGTGTAGAGCTTTT 1048  
QY 367 TATTAACATCAATCTTTAAGACATCATATATCTGTT---CATGAAAAACATTAACG 423  
DB 1049 AACCAACCTCAAGCTTACTGTAACATTAATTAATCATCCGAGAGAAATGCTACAAA 1108  
QY 424 TGTAAACATGTAATAAGTTTTCACCTGACCTTCAAAATTAAGACAACTTAATTAA 483  
DB 1109 TGTGAAGAAATGTGGCAAAAGTTTAAACCGATCTTCAAACTTACAGACATTAATAATTT 1168  
QY 484 CATCATGTG 493  
DB 1169 CATTAAGAG 1178

RESULT 5  
ID ACAS6456 standard; cDNA; 2320 BP.  
XX  
AC ACAS6456;  
XX  
DT 06-JUN-2003 (first entry)  
XX  
DE Human signalling pathway polynucleotide probe SEQ ID NO 1054.  
XX  
KW Human; probe; ss; array element; Parkinson's disease;  
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;  
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.  
OS Homo sapiens.  
XX  
PN US6500938-B1.  
XX  
PD 31-DEC-2002.  
XX  
PF 30-JAN-1998; 98US-00016434.  
XX  
PR 30-JAN-1998; 98US-00016434.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Au-Young J, Seilhamer JJ;  
XX  
DR WPI; 2003-352189/33.  
XX  
PT Combination of polynucleotide probes, useful as array elements in a  
PT microarray for monitoring the expression of a number of target  
PT polynucleotides.  
XX  
PS Claim 1, SEQ ID NO 1054; 65pp; English.  
XX  
CC The invention relates to a combination which, comprises a number of  
CC polynucleotide probes comprising a sequence selected from one of the 1490  
CC sequences mentioned in the specification. The combination is useful as an  
CC array element in a microarray for monitoring the expression of a number  
CC of target polynucleotides. The microarray is particularly useful in the  
CC diagnosis and treatment of cancer and immunopathology and neuropathology.  
CC The microarray is useful in diagnostic and treatment regimens, drug  
CC discovery and development, toxicological and carcinogenicity studies.

CC forensics and pharmacogenomics. The microarray is also useful for  
CC monitoring progression of diseases and for developing sophisticated  
CC profiles for the effects of currently available therapeutic drugs. The  
CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
CC and genomic fragments and in research and diagnostic applications. The  
CC array can detect changes in expression in a large number of genes coding  
CC for different signalling pathway populations which can be used to diagnose  
CC various diseases including cancer e.g. adenocarcinoma and leukaemia,  
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease  
CC and Parkinson's disease. The present sequence represents a polynucleotide  
CC probe of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=06500938B1  
XX

Seq Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;  
Query Match 7.8%; Score 96.6; DB 10; Length 2320;  
Best Local Similarity 51.3%; Pred. No. 1.7e-09;  
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGTATATTTGACATATGAAGGTTGTATTAAGCTATATTCGACATCATATTATTA 128  
DB 964 AGAAGAAACCTTACAAATGTAAAGATGTGTAAAGCTTTTAAACGATCTTCAACCTTAC 1023  
QY 129 GCAACATTTAAGAACCCAGATATGATCGACCGTATTAATGTATCAAGTGGACATTTGTA 188  
DB 1024 TACCCATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGATGTGG 1077  
QY 189 TAAAGCATTTTTCAGAAATTCACATTTGGAACACATATTTGATCATATTCGAAAAAA 248  
DB 1078 CAAGGCTTTAAGCAGTCCCAAACTTACTACACATTAAGATTAATTCATATCTGAGAGAA 1137  
QY 249 ACCATTTCATTTTCAGTGTGTGTAAAGGGTTAATTCGACAAACATTTGAAAAACA 308  
DB 1138 ACCCTCAAAATGTAAAGATGTGAAAGAGCTTTTAAACAGTCTGCACACCTTACACACA 1197  
QY 309 TGAATCACCCATCAAAAGTCATTTAAATGTATCATTTGAAATTTGCAAGACATTTTA 368  
DB 1198 TGAAGTATTCATATCTGAGAGAAACCTTCAAAATGTGAAAGAGCTTTAA 1257  
QY 369 TAAACA---TCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAAAACATTTA 425  
DB 1258 TCATTTCTCACACCTTACATCAATTAAGATTAATCTACTGAGAGAAACCTTAA 1317  
QY 426 TAAACATGTATTAAGTTTTCCTGACCTTCAAAATTAAGCAACATTAATTAACA 485  
DB 1318 TAAAGATGTGTAAAGCTTTTAAACATCTTCAACCTTACATTAAGATTAATTA 1377  
QY 486 TCATGTGATCTCCCTGCTTATCAATGTATCATCTCGTGTGTTTAAAAATTTCCAAAC 545  
DB 1378 TACTGAGAGAGAACCTTACAAATCTAAAGATGTGAAAAAGCTTTTAAACCAATCTCA 1437  
QY 546 TTGTCAGTATTAATTAATTAATTAACACATGTCATCCAAATTAATGTCCTTAATG 605  
DB 1438 ACTTACTGAACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTAAAAAATG 1494  
QY 606 TGGTAAAGTTGTGTGGGAAAAAGGTTTATCTTCAATATGTTAATGTCATGTA 662  
DB 1495 TGGCAAGCTTTTAAACAGTCTCAAAATCTTACTAGACATTAAGAAAGTCATACAG 1551

RESULT 6  
ID ADI56252 standard; DNA; 2320 BP.  
XX  
AC ADI56252;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human polynucleotide probe #1054.  
XX  
KW Human; probe; ss; receptor-like polypeptide; transducing polypeptide;

KW	effectorlike polypeptide; cancer; immunopathology; neuropathology;
KW	drug development; toxicology; carcinogenicity;
KW	signalling pathway polypeptide; adrenal gland; bladder; bone;
KW	bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;
KW	diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;
KW	dementia; amnesia; epilepsy; Alzheimer's disease; depression.
XX	
OS	Homo sapiens.
XX	
PN	US2004010136-A1.
XX	
PD	15-JAN-2004.
XX	
XX	26-NOV-2002; 2002US-00305720.
XX	
PR	30-JAN-1998; 98US-00016434.
XX	
XX	(INCY-) INCYTE GENOMICS INC.
PA	
XX	
PI	Au-Young J, Selthamer JU;
XX	
DR	WPI; 2004-090520/09.
XX	
PT	New composition comprising polynucleotide probes, useful as array
PT	elements in a microarray for monitoring the expression of target
PT	polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic
PT	fragments.
XX	
PS	Claim 6, SEQ ID NO 1054, 73bp; English.

	Query Match	7.8%	Score 96.6	DB 12	Length 2320
	Best Local Similarity	51.3%	Pred. No. 1.7e-09		
	Matches 306	Conservative	0	Mismatches 279	Indels 12
				Gaps	3
Qy	69	AAAGTATTTTGCACATATGAAAGGTGTGATAAAGCCCATATATGCACCATCATTTATTTAGA	128		
Db	964	AGAGAAACCCATCAAAATGTTAAAGATGTGTAAAGCTTTTAAACGATCTTCAACCCTTAC	1023		
Qy	129	GCAACATTTTAAAGAACCCACAGTAATGATGCAACGTAAATGTACATGACGATTTGTGA	188		
Db	1024	TACCCATGAAAAAATTCTACTGAGAGAAACCTTACAAATGT-----GAAGATGTGG	1077		
Qy	189	TAAAGCATTTTTCAGAAAAATCACATTTTGGAAACACATATTGTATTCACATTTCCGAAAAAA	248		
Db	1078	CAAGGCCCTTTAAGCAGTCCCTCAAAACCTTACTACACATTAAGATTAATTTCTACTGTGAGAGA	1137		

QY	249	ACCAATTCATGTTCAAGTGTGGTTAAAGSGTTAAATTCTCGACAAACCTTGAAGAAC	308
Db	1138	ACCCACAAATGTAAAAAATGTGGAAAGCCTTTAACCAAGTCTGCACACCTTACACACA	1197
QY	309	TGAATCACCCATACAAAGTCATTAAATGTACATTGAAATTTGTACAGAGCATTTTA	368
Db	1198	TGAGGTATTATATCTGGAGAGAAACCCCTCAATATGTGAAAAAATGTGGAAAGCCTTTAA	1251
QY	369	TAAACA--TCAATCTTTAAGACATCATATTTATCTGTTCATGAAAAAACAATTAAAGTG	425
Db	1258	TCATTTCTCACACCTTACTACACATTAAGATTAATCAATCGAGAGAAACCTTCAAAATG	1317
QY	426	TAAACATGTATTAAGTTTCACTGCACCTTCAAAATTTAGCACACATTAATTAAAAACA	485
Db	1318	TAAAGAAATGTGTAAAGCTTTTAAACCTTCAACCCCTTACAAACATTAAGATTAATTCA	1377
QY	486	TCATGTGTGAAATCTCTGCTTATCAATGTGATCATCTGTGTGTTTAAAAATTTCCAAC	545
Db	1378	TACTGTGAGAGAGCCTTACAAATCTTAAGAAATGTGAAAAAGCTTTTAAACCATCTCTCAA	1437
QY	546	TTGGTCAGTATTACAATTTCAATATTAACAACATGCAATCCAAAAGTTAAATGTCTTAATG	605
Db	1438	ACTTACTGAACATTAAGAAAATTTCACTACTGAGAGAAACCCCTA---TGAATGTGAAAAATG	1494
QY	606	TGTGTAAGGTGTGTGGGAAAAAAGTTTATCTTCAATATGTGTAAGTATGATGA	662
Db	1495	TGGCAAGACCTTTTAAACAGTCTTCCAAATCTTATACATGACATTAAGAAAAGTCAATACAGA	1551

RESULT 7  
 AAS91317  
 ID AAS91317 standard; cDNA; 4563 BP.  
 AC AAS91317;  
 XX  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 XX  
 DE DNA encoding novel human diagnostic protein #27121.  
 XX  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF  
 XX 31-MAR-2000; 2000US-00540217.  
 PR  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR  
 DR P-PSDB; ABG27130.  
 XX  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PT  
 PS Claim 1; SEQ ID NO 27121, 103pp; English.  
 XX  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal



CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. Aa664197-Aa694564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;

XX Query Match 7.7%; Score 95.2; DB 5; Length 4563;

XX Best Local Similarity 52.7%; Pred. No. 3.5e-09;

Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

QY 7 GAAAGTACGAAACCAATGATATCATCTTAAATCTTCTTCTATCAGCGTCC 66

Db 1315 GAATGGGAAAGCCTTTAACAGTCTCTACACCTTACTCAACATTAACATTCATAC 1374

QY 67 AAAAAGTATTTTGGCATTATGAAAGGTGTGATTAAGCTTATATGACCATATTATTA 126

Db 1375 GGAGGAAAAACCTACAAATGTGAAGATGTGGCAAGCTTTTAACTTTTCAAGCCCTT 1434

QY 127 GACACATTTTAAGAACCCACATTAATGATGACCGTATAATGTAAGTGAAGCATTTG 186

Db 1435 ACTTAACATTAAGTAAATTTCTCTGGGAGAAACCAATCAATGT-----GAAGATGT 1488

QY 187 GATTAAGCATTTTTCAGAAATGACATTTGAAACATATTGTATCATCTCCGAAAA 246

Db 1489 GGGAAAGCTTTTGGCAATCTCTACACCTTACTAGCATTAAGCAATTCATCTGAGAG 1548

QY 247 AAAACATTTCCATTTGTCAGTGTGTGTGAAGGGGTTAATTCGACAAACCTTGAAAA 306

Db 1549 AAACCTCAAAATGTGAAGATGTGGCAAGCTTTTAACTTTCTCAGACCTTAGAAG 1608

QY 307 CATTGAATGCCATTAAGTCAATTTAATGTACATTTGAATGTAAGTGAAGCATTT 366

Db 1609 CATTAAGTAAATTTACTGAAAGAAACCTTAACAAATGTGAAGATGTGGAAAGCTTTT 1668

QY 367 TATAAATCATCAATCTTAAAG---CATCATATATTTATCTGTCTATGAAAAAACATTAA 423

Db 1669 AGCGAGTCTCAACCTTAAAGAACATCAGATTAATTCATCTGAGAGAAACCTTAACA 1728

QY 424 TGTAAACATGTATAAAGTTTTCATCTGACCTTCAAAATTTAGACAACTAAATTAAAA 483

Db 1729 TGTGAAGATGTGTGAAGCTTTTAAGTGTTCATCAAACTTCTGATCAATTAAGTAA 1788

QY 484 CATCAGGTG 493

Db 1789 CATACGTGAG 1798

RESULT 8

ID ADN04717 standard; cDNA; 981 BP.

XX ADN04717;

DT 01-JUL-2004 (first entry)

XX Antipsoriatic cDNA sequence #572.

XX des; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.

XX Homo sapiens.

XX

PN WO2004028479-A2.

XX PD 08-APR-2004.

XX PF 25-SEP-2003; 2003WO-US030907.

XX PR 25-SEP-2002; 2002US-0414006P.

XX (GENTH) GENENTECH INC.

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WJ,

XX Wu TD;

XX WPI; 2004-305105/28.

XX P-PSDB; ADN04718.

XX New PRO nucleic acid or polypeptide, useful for preparing a

XX pharmaceutical composition for diagnosing or treating psoriasis in a

XX mammal.

XX Claim 1; SEQ ID NO 1111; 3069bp; English.

XX The invention relates to novel polynucleotide and polypeptides for

XX treating psoriasis or a sequence having at least 80% identity to the

XX above sequences. The nucleic acid is useful for preparing a composition

XX for diagnosing or treating psoriasis in a mammal. This sequence

XX corresponds to one of the polynucleotides of the invention.

XX Sequence 981 BP; 380 A; 178 C; 170 G; 253 T; 0 U; 0 Other;

XX Query Match 7.7%; Score 95; DB 12; Length 981;

XX Best Local Similarity 50.4%; Pred. No. 3.1e-09;

XX Matches 289; Conservative 0; Mismatches 275; Indels 9; Gaps 2;

QY 89 AAGGCTGTATTAAGCTTAATTCGACATCATTTATTAGACAACTTAAGAACCCACA 148

Db 1 AAGAAATGTGCAAAAGCTTTAGATATTTCTCAACCTTACTTAACATTAAGATTCATTA 60

QY 149 GTATATGATGACCGTATTAATGTAAGTGAAGATGTGTGATTAAGATTTTCAAGAAAT 208

Db 61 CTGGAGAGAAACCTTAACAAATGCA-----ATGAATGTGTAAAGCTTTAACTGTCTCT 114

QY 209 CACATTTGAAACACATATTGTATCATCTCCGAAAAAAACCATTTCCATTTGTCAGTGT 268

Db 115 CAACCTTTACTTAACATTAAGATTCATCTGAGAGAAACCTTAACAAATGTGAAGAAAT 174

QY 269 GTGGTAAAGGGGTTAATTTCTGACCAACCTGAAAAAGACATGAATCACCATTAACAAGT 328

Db 175 GTGGCAAAAGCTTTTAAACAGTCTCAACCTTAAGATTAAGATTAAGTCAATCTGAG 234

QY 329 CATTTAATGTACATTTGAATTTGCAAGAGCATTTTATTAACATCAATCTTTAAGC 388

Db 235 AGAAACCTTAACAAATGTGAAGATGTGTAAAGCTTTTAAACGTCACAACTTAAGC 294

QY 389 ATCATATATTTATCTGTCAGAAAAAATTAACGTTAAACGTTAAATTAAGTTTCA 448

Db 295 ATTAAGAAATTTATTAAGAAACCAATCAATGTGAAGATGTGAAGAAAGCTTTA 354

QY 449 CTGACCTTCAAAATTTAGCAACATTAATTAAGATGATGATCTCTGCTTATC 508

Db 355 GTGTATTTCAACCTTACTTAAGATTAAGTATTTATCTGAGCAAAACCTTAACAAAT 414

QY 509 AATGTATCATCTCTGTTGTTTAAATTTTCAAACTTGCTGATGATTAATTTCAAT 568

Db 415 GTGACGAATGTGCAAGTGTCTTAAAGGCAATTCACACCTTACTGAACATTAAGAGTTC 474

QY 569 TAAACAACTGCAATCCAAAACTTAATGCTTAATGTGTGAAGTGTGTTGGGAAA 628

Db 475 ATACTGAGAGAAACCTTAAC---AAATGCAATGAATGTGTAAAGCTTTAACTGATCT 531

QY 629 AAGTTTATCTTCATATGATTTAAGTCATGATG 661

Db 532 CAACCTTACTTAACATTAAGAGATTCATACCTG 564

RESULT 9  
AAL51569

ID AAL51569 standard; DNA; 2597 BP.

XX AAL51569;

DT 10-APR-2003 (first entry)

DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.

XX Human; gene; de; nucleic acid-associated protein; NAAP; arteriosclerosis;  
XX cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;  
XX cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;  
XX mental retardation; neurological disorder; Alzheimer's disease; epilepsy;  
XX Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;  
XX Crohn's disease; transgenic animal; animal model.

OS Homo sapiens.

PN WO200300864-A2.

PD 03-JAN-2003.

PF 20-JUN-2002; 2002WO-US021179.

PR 22-JUN-2001; 2001US-0300518P.

PR 29-JUN-2001; 2001US-0301787P.

PR 29-JUN-2001; 2001US-0301792P.

PR 29-JUN-2001; 2001US-0301892P.

PR 06-JUL-2001; 2001US-0303405P.

PR 06-JUL-2001; 2001US-0303442P.

PR 15-MAR-2002; 2002US-0364438P.

XX (INCT-) INCTE GENOMICS INC.

XX Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;  
XX Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IU;  
XX Ramkumar J, Griffin JA, Yang J, Sanjanwala NM, Baughn MR;  
XX Borowsky ML, Yao MG, Wallia NK, Bandman O, Lai PG, Becha SD, Lee SY;  
XX Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjian Y, Lu Y;  
XX WPI; 2003-201420/19.  
XX P-PSDB; AAO16419.

PT New nucleic acid-associated proteins and polynucleotides, useful for  
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),  
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune  
PT disorders (e.g. AIDS).

PS Claim 12; Page 292-293; 312pp; English.

XX The invention comprises the amino acid and coding sequences of human  
XX nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
XX the invention are useful for diagnosing, treating or preventing disorders  
XX associated with aberrant expression of NAAP, such as: cell proliferative  
XX disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis  
XX or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia  
XX or mental retardation); neurological disorders (e.g. Alzheimer's disease,  
XX Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders  
XX (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of  
XX the invention are useful for creating transgenic animals to model human  
XX disease. The present DNA sequence encodes a human nucleic acid-associated  
XX protein of the invention

XX Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

XX Query Match 7.7%; Score 94.6; DB 8; Length 2597;

XX Best Local Similarity 55.1%; Pred. No. 4.2e-09;  
XX Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;

QY 78 TTGACATATGAGGGGTGTGATTAAGCCTATATATGACCATTATTAAGCAACATT 137  
DB 1070 TTATTAATGTGAAGATGTGGCAAGCCTTATATGTCTCAACCTTACTCAACATA 1129  
QY 138 AAGAACCCACAGTATATGATCGACCGGTATTAATATGTACAGTGCATTTTGATTAAGCAT 197  
DB 1130 GAGAAATTCACTACGAGAGAAACCTTACAAATGT-----GAAAGGTGGCAAAAGCCTT 1183  
QY 198 TTTCAAAAAATCACAATTGGAACACATATTGTATCATCTTCCGAAAAAAACCATTCGA 257  
DB 1184 TTAAGTGTCTTCACTCTTCACTCAACATTAAGAAATTCATCTGAGAGAAAAACCATCA 1243  
QY 258 TTGTTCAAGTGTGTGTAAGGGGTATATCTTCGACAACATTTGAAGAATCAATCA 317  
DB 1244 ATGTAAGAATGTGGCAAGCCTTTAACAATCTTCACACCTCACACATATAAAGAT 1303  
QY 318 CCATCAAAAGTCATTTAATATGATTAATTTGAATTTGTCAAGAGATTTTATAA--ACA 374  
DB 1304 TCATACCGGAGAGAAACCTTACAAATGTGAAGAAATGTGGCAAAAGCCTTTAACCACTTC 1363  
QY 375 TCATCTTTAAGACATCATATATATTTCTGTCATGAAAAAATTAAGGTAAACAT 434  
DB 1364 ACACTTACTACATATATATATATATATCTGAGAGAAACCTTCAAAATGTAAAGAT 1423  
QY 435 TATATAAGTTTCACTGCACTTCAAAATTAAGCAACATATAATTAACATCATGTTG 493  
DB 1424 TGGCAAGCTTTTAAGCGGTCTCTCAACCTTACTGAACATATGATTAATTCATATCGAG 1482

RESULT 10

AAS64586  
ID AAS64586 standard; CDNA; 3639 BP.

XX AAS64586;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #390.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HSE-) HSE INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABC00399.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostic, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 1; SEQ ID NO 390; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

Query Match 7.5%; Score 92.4; DB 5; Length 3639;  
 Best Local Similarity 54.1%; Pred. No. 1.2e-08;

Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

QY 69 AAGTATATTGACATATGAGGGTGTGATTAAGCTTATATGACCATCATATTATTAGA 128  
 DB 738 AGGAAACCCCTACAAATGTGAGAAATGTGGCAAAAGCTTTAGCCATCTTCAACCTTGC 797  
 QY 129 GCAACATTTAAGAACCCAGTATATGATCGACCGTAAATGTAAGTGAAGCATTTGCA 188  
 DB 798 TAAACATAGAGAAATTCATCTGAGAGAAACCCCTAAATGTG-----GAAAGATGTGG 851  
 QY 189 TAAAGATTTTTCAGAAATACATTTTGAAGACATTTGTATCATATTCGAAAAAAA 248  
 DB 852 CAAGGCTTTAGCGCTTCTCAACCCCTTGTAAACATTAAGAAATTCATCTGAGAGAA 911  
 QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTATTTCTGACAAACATTTGAAGAGCA 308  
 DB 912 ACCCTCAATGTAAAGATGTGGCAAAAGCTTTAGCAATCTTCAACCTTGTCTATCA 971  
 QY 309 TGAATACCCCATACAAAGCATTTAATATGATTTGAAGAAATTTGCAAGAGCATTTTA 368  
 DB 972 TAAGATTAATCAATCTGAGAGAAACCCCTCAAAATGTAAGAAATGTGACAAACCTTTTA 1031  
 QY 369 TAAACATCAATC---TTTAGACATCATATATTATCTGTCATGAAAAAATTTAAGTGG 425  
 DB 1032 GGCATCTCAACCCCTTACTTAACATTAATTAATATCATCTGAGAGAAACCTTACAAATG 1091  
 QY 426 TAAACATGTAATTAAGTTTTCATCTGACCTTCAAAATTAGCAACAATTAATTAAGCA 485  
 DB 1092 TGAAGAAATGTGGCAAGCTTTTATGATCTTCAAAATCTTACTATATATAGTTTATCA 1151  
 QY 486 TCATGTGTGATCTTCT 501  
 DB 1152 TACTGAGAGAAACCT 1167

RESULT 11

ACH87568/C  
 ID ACH87568 standard; DNA; 994 BP.

XX ACH87568;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX Human genome derived single exon probe #20763.  
 XX  
 XX Human; probe; ss; gene expression; single exon probe; microarray;  
 KM alternative splicing event; genomic alteration.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2003194704-A1.  
 PN  
 XX 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.  
 PF  
 XX  
 PR 03-APR-2002; 2002US-00029386.  
 XX  
 XX (PENN/) PENN S G.  
 PA (RANK/) RANK D R.  
 PA (HANZ/) HANZEL D K.  
 XX  
 PI Penn SG, Rank DR, Hanzel DK;  
 XX  
 DR WPI; 2004-119264/12.  
 XX

PT New human genome-derived single exon nucleic acid probes useful for human  
 PT gene expression analysis, for identifying or characterizing alternative  
 PT splicing events, for assessing genomic alterations or as tools for  
 PT surveying tissues.  
 PT  
 XX

PS Claim 1; SEQ ID NO 20763; 80bp; English.

CC The invention relates to a nucleic acid probe for measuring human gene  
 CC expression, comprising any of the 27,400 fully defined nucleotide  
 CC sequences in the specification, or their complements or fragments, and  
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences  
 CC fully defined in the specification. The probe is a single exon probe that  
 CC hybridizes under high stringency conditions to a nucleic acid molecule  
 CC expressed in human cells or tissues. Also included are a spatially-  
 CC addressable set of single exon nucleic acid probes for measuring human  
 CC gene expression (comprising a plurality of single exon nucleic acid  
 CC probes cited above, where each of the plurality of probes is separately  
 CC and addressably isolatable or amplifiable from the plurality), a single  
 CC exon microarray for measuring human gene expression, a method of  
 CC measuring human gene expression, a vector comprising the single exon  
 CC probe cited above, an ORF-encoded peptide comprising at least 8  
 CC contiguous amino acids of any of the above-mentioned amino acid  
 CC sequences (optionally with conservative amino acid substitutions), an  
 CC isolated antibody that binds specifically to a peptide cited above,  
 CC methods of selling and/or licensing single exon probes or microarrays to  
 CC a customer desiring to measure gene expression, a method of providing  
 CC human gene expression data by subscription, and a computer-readable  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above. The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterizing  
 CC alternative splicing events, in detecting and characterizing gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docid=20030194704  
 XX  
 XX

Sequence 994 BP; 252 A; 159 C; 203 G; 380 T; 0 U; 0 Other;

Query Match 7.4%; Score 92; DB 12; Length 994;  
 Best Local Similarity 52.0%; Pred. No. 1.2e-08;

Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 7 GAAAGTGACGAACCAATGATATCATTTTAATATCTTCTTCTTCATCAGTCCC 66  
 DB 763 GAATGTGGCAAGCCTTCTACCATTTCTCAACCTTACTACATTAAGTAATTCATCTACT 704  
 QY 67 AAAAGTATATTGACATATGAGGGTGTGATTAAGCCTTATTAATGACACATATTAATTA 126  
 DB 703 GAGAGAAAGCCCTTCAAAATGTGAAGATGTGTAAGCTTTTAAACACCCCTTCAAGCCCTT 644  
 QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGAAGATTTG 186  
 DB 643 ACTACACATTAATGTCTTATGTTTAAGAAAAACCTTACAAATGT-----GAAAGATGT 590

QY 187 GATAAGCATTTTTCAGAAAATCACATTGGAAAACATATGTATCACATTCGGAATAA 246  
 DB 589 GACAAAGCTTTAAACGATTCATCTACTTAACATTAAGTAATTCATCTGGAGAG 530  
 QY 247 AAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCCTGCACAACTTGAATAA 306  
 DB 529 AAATCTTAACAAATGTGAACAAATGTGGCAAGGGTTTAAGTGTCTTCAACCTTACAAA 470  
 QY 307 CATGAATCACCCATACAAAGTCATTAAATGATGTAATGTAATGTAATGTAATGTAATG 366  
 DB 469 CATTAAGAAATTCATCTCTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGGCTTT 410  
 QY 367 TATTAACAT---CAATCTTTAAGACATCATATATATCTGTCATGAAAAAATTTACG 423  
 DB 409 AATGTGTCTTACACCTTCTACATTAAGATGATTCATACGAGAGAAACCTTACAAA 350  
 QY 424 TGTAAACATGTATTAATGAATTTTCACTCGACCTTCAAAATTTAGCACAATTAATTA 483  
 DB 349 TGTGAAGAAATGTGGCAAGGCTTTAACCCTCTCAAACTTACTATATGAATTAAT 290  
 QY 484 CATCATGTGTGATCTTCT 501  
 DB 289 CATCTGGAGAGAAACCT 272  
 RESULT 12  
 ADRI4368  
 ID ADRI4368 standard; DNA; 2873 BP.  
 AC ADRI4368;  
 XX  
 DT 21-OCT-2004 (first entry)  
 XX  
 DE Human NF-kappaB pathway-associated gene SeqID369.  
 XX  
 KW NF-kappaB pathway; antiinflammatory; cytoskeletal; hepatotropic; virucide;  
 KW antiatheritic; antirheumatic; gastrointestinal-gen; antiaesthetic;  
 KW antiarteriosclerotic; immunomodulatory; cerebroprotective; vasotropic;  
 KW immunosuppressive; vulnery; gene therapy; immune disorder;  
 KW inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;  
 KW hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;  
 KW hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;  
 KW X-linked amideotic ectodermal dysplasia; immunodeficiency;  
 KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;  
 KW viral replication; host cell survival; evasion of immune response;  
 KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;  
 KW autoimmunity; cachexia; euthyroid sick syndrome; stroke; EAE;  
 KW aberrant acute phase response; hypercongenital condition; birth defect;  
 KW necrotic lesion; wound; organ transplant rejection;  
 KW aberrant signal transduction; proliferating disorder; cancer;  
 KW HIV propagation; gene; de; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004065577-A2.  
 XX  
 PD 05-AUG-2004.  
 XX  
 PF 13-JAN-2004; 2004WO-US000798.  
 XX  
 PR 14-JAN-2003; 2003US-0440068P.  
 XX  
 PR 12-MAY-2003; 2003US-0469757P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Nadler SG, Neubauer MG, Feder JN, Carman J;  
 XX  
 DR WPI, 2004-562168/54.  
 XX  
 DR P-PSDB; ADRI4369.  
 XX  
 PT New isolated polynucleotides and polypeptides associated with NF-kappaB

PT pathway, useful for diagnosing, treating, or preventing disorders or  
 PT diseases associated with NF-kappaB pathway.  
 XX  
 PS Claim 1; SEQ ID NO 369; 237bp; English.  
 XX  
 CC This invention relates to the novel association of protein sequences (and  
 CC the genes which encode them) to the NF-kappaB pathway. The invention may  
 CC be useful for the production of compounds with an antiinflammatory,  
 CC cytoskeletal, hepatotropic, virucide, antiaesthetic, antirheumatic,  
 CC gastrointestinal-gen, antiasthmatic, antiarteriosclerotic,  
 CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or  
 CC vulnery activity or for gene therapy. The proteins and nucleotides are  
 CC useful for diagnosing, preventing, treating, or ameliorating conditions  
 CC or diseases associated with the NF-kappaB pathway. The condition is an  
 CC immune disorder, an inflammatory disorder, an inflammatory disorder  
 CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,  
 CC hepatic disorders, Hodgkin's lymphoma, haematopoietic tumours, hyper-IgM  
 CC syndromes, hypohidrotic ectodermal dysplasia, X-linked amideotic  
 CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,  
 CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell  
 CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory  
 CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick  
 CC syndrome, stroke, EAE, autoimmunity disorders, disorders related to hyper  
 CC immune activity, disorders related to aberrant acute phase response,  
 CC hypercongenital conditions, birth defects, necrotic lesions, wounds,  
 CC organ transplant rejection, conditions related to organ transplant  
 CC rejection, disorders related to aberrant signal transduction,  
 CC proliferating disorders, cancers and HIV propagation in cells infected  
 CC with other viruses. The present sequence is that of a human gene which is  
 CC subject to the novel association with the NF-kappaB pathway of the  
 CC invention. Note: This sequence does not appear in the specification but  
 CC was obtained by the indexer from Genbank.  
 XX  
 SQ Sequence 2873 BP; 948 A; 564 C; 545 G; 816 T; 0 U; 0 Other;  
 Query Match 7.4%; Score 92; DB 13; Length 2873;  
 Best Local Similarity 52.2%; Pred. No. 1.4e-08;  
 Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;  
 QY 7 GAAAGTGAAGAAACCAATTCATATCTTAAATCTCTTCTTCATCAGTCC 66  
 DB 1621 GAATGTGCGAAACATTTTTCATGCTTTCACCTTACCAATTAAGATTCAAACT 1680  
 QY 67 AAAAGTATTTTGGACATATGAAGGTTGTAAAGCCTTAATGACATCATTTATTA 126  
 DB 1681 AGAGGATTTTCTACAAATGTGAAGCATATGAAGAGCCTTAACTGTGCTCAACCTT 1740  
 QY 127 GAGCAACATTTAAGAACCCAGTATGATGACCGGTAAATTTAGTACGTGACGATGT 186  
 DB 1741 AATTAACATTAAGAGATTCATCTGAGAGAAACCTTCAATGTGAAA-----GAATGT 1794  
 QY 187 GATTAAGCATTTTTCAGAAAATCACATTGGAAAACATATGTATTCATTCGGAATAA 246  
 DB 1795 GGCAAAAGCCTTTAACCAGACCTTATTAACATTAAGAGATTCATCTGAGAGAG 1854  
 QY 247 AAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCCTGCACAACTTGAATAA 306  
 DB 1855 AAACCTTAACAAATGTGAACAAATGTGGCAAGGGCTTTAACAGTCAATCGACCTTACTACA 1914  
 QY 307 CATGAATCACCCATACAAAGTCATTAAATGATGTAATGTAATGTAATGTAATGTAATG 366  
 DB 1915 CATATATATATTCATATCGGGGAAATTCCTCAATGTGAAGAAATGTGTAGACTTTT 1974  
 QY 367 TATTAACATCAATCTTTAAGACATATATATCTGTT---CATGAAAAAATTTACG 423  
 DB 1975 AACCAAGCCTCAAGCCTTACTGAACATTAATTAATTCACCGGAGAGAGAGTTATGAA 2034  
 QY 424 TGTAAACATGTATTAATGAATTTTCACTCGACCTTCAAAATTTAGCACAATTAATTA 483  
 DB 2035 TGTGAAGAAATGTGGCAAGGCTTTTAAACGATCTCAAACTTACTGAGACATTAAGTACATT 2094  
 QY 484 CATCATGTGTG 493



```

Db      659 AATTCTCGATTCCTTATATAACATTAAGAGATTTCATATGAGATTAACCTTCAATGTG 718
Oy      428 AACAAATGATATAAGTTTCACCTGACCTTCAAAATTTAGCACACATAATTAACATC 487
Db      719 AAGATGTGGCAAGCGCTTATAGATGTCTTCTCAATTTCTTAAAAACATAAGATATACATA 778
Oy      488 ATGTGTGATCTCC 500
Db      779 CTGGGAAAAACC 791

RESULT 14
AD000633
ID      AD000633 standard; cDNA; 1533 BP.
XX
XX
AC      AD000633;
XX
DT      29-JUL-2004 (first entry)
XX
DE      Novel human cDNA sequence #1448.
XX
KW      de; gene; anti-inflammatory; dermatological; neuroprotective;
KW      immunomodulator; antibacterial; virucide; antiparasitic; cytostatic;
KW      gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW      psoriasis; diabetes; early aging; hormonal imbalance;
KW      ischemic heart disease; ulcerative colitis.
XX
XX      Homo sapiens.
XX
XX      WO2004038003-A2.
XX
PD      06-MAY-2004.
XX
PF      24-OCT-2003; 2003WO-US033947.
XX
XX
PR      25-OCT-2002; 2002US-0421061P.
PR      25-OCT-2002; 2002US-0421080P.
PR      25-OCT-2002; 2002US-0421552P.
PR      25-OCT-2002; 2002US-0421614P.
PR      30-OCT-2002; 2002US-0422177P.
PR      30-OCT-2002; 2002US-0422178P.
PR      15-NOV-2002; 2002US-0426355P.
PR      15-NOV-2002; 2002US-0426384P.
PR      15-NOV-2002; 2002US-0426394P.
PR      15-NOV-2002; 2002US-0426430P.
PR      15-NOV-2002; 2002US-0426916P.
PR      27-NOV-2002; 2002US-0429244P.
PR      27-NOV-2002; 2002US-0429275P.
PR      27-NOV-2002; 2002US-0429302P.
PR      27-NOV-2002; 2002US-0429326P.
PR      27-NOV-2002; 2002US-0429651P.
PR      04-DEC-2002; 2002US-0430645P.
PR      04-DEC-2002; 2002US-0430651P.
PR      04-DEC-2002; 2002US-0430657P.
PR      04-DEC-2002; 2002US-0430663P.
PR      04-DEC-2002; 2002US-0430668P.
PR      04-DEC-2002; 2002US-0430684P.
PR      05-DEC-2002; 2002US-0430937P.
PR      05-DEC-2002; 2002US-0430965P.
PR      05-DEC-2002; 2002US-0431458P.
PR      12-DEC-2002; 2002US-0433251P.
PR      12-DEC-2002; 2002US-0433500P.
PR      13-DEC-2002; 2002US-0433316P.
PR      13-DEC-2002; 2002US-0433318P.
PR      23-DEC-2002; 2002US-0436236P.
PR      03-JAN-2003; 2003US-0437914P.
PR      17-JAN-2003; 2003US-0440820P.
PR      17-JAN-2003; 2003US-0440821P.
PR      18-APR-2003; 2003US-0463700P.
PR      18-APR-2003; 2003US-0463708P.
PR      18-APR-2003; 2003US-0463716P.
PR      18-APR-2003; 2003US-0463732P.

```

```

PR      02-MAY-2003; 2003US-0467199P.
PR      02-MAY-2003; 2003US-0467201P.
PR      02-MAY-2003; 2003US-0467203P.
PR      02-MAY-2003; 2003US-0467203P.
PR      02-MAY-2003; 2003US-0467203P.
PR      19-MAY-2003; 2003US-0471306P.
PR      19-MAY-2003; 2003US-0471336P.
PR      22-MAY-2003; 2003US-0472420P.
PR      22-MAY-2003; 2003US-0472430P.
PR      09-JUN-2003; 2003US-0476609P.
PR      09-JUN-2003; 2003US-0476621P.
PR      09-JUN-2003; 2003US-0476632P.
PR      09-JUN-2003; 2003US-0476641P.
PR      08-JUL-2003; 2003US-0485217P.
PR      08-JUL-2003; 2003US-0485218P.
PR      08-JUL-2003; 2003US-0485223P.
PR      08-JUL-2003; 2003US-0485224P.
PR      08-JUL-2003; 2003US-0485325P.
PR      08-JUL-2003; 2003US-0485359P.
PR      14-JUL-2003; 2003US-0486446P.
PR      14-JUL-2003; 2003US-0486480P.
PR      15-JUL-2003; 2003US-0486891P.
PR      15-JUL-2003; 2003US-0486960P.
PR      08-AUG-2003; 2003US-0493341P.
PR      08-AUG-2003; 2003US-0493370P.
PR      08-AUG-2003; 2003US-0493573P.
PR      08-AUG-2003; 2003US-0493577P.

(FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
XX      William LT, Chu K, Lee E, Heistr K, Beaurang PA, Behrens D;
XX      Halenbeck RF, Kochakota S, Lin H, Linemann T, Pierce K, Wang Y;
XX      Wong JG, Wu G, Zhang H, Zeng C;
XX
XX      WPI; 2004-365511/34.
XX
XX      P-PSDB; ADM99848.
XX
PT      New nucleic acid molecules, useful in preparing a composition for
PT      treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT      disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT      ulcerative colitis.
XX
XX
PS      Claim 1, SEQ ID NO 2232; 532pp; English.
XX
XX
XX      The invention relates to a nucleic acid molecule comprising a
XX      polynucleotide sequence or its complement that encodes a polypeptide. The
XX      nucleic acid is useful in preparing a composition for treating or
XX      preventing inflammatory, CNS, immune, bacterial or viral disorder,
XX      cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
XX      heart disease or ulcerative colitis. This sequence corresponds to a
XX      nucleic acid of the invention.
XX
XX
SQ      Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;
XX
Query Match      7.4%; Score 91.8; DB 12; Length 1533;
Best Local Similarity 52.1%; Pred. No. 1.4e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;
Oy      11 GTGAGAAACCAATGATGATCATCTTTAATATCTCTTCTTCATCAGTCCCAAA 70
Db      305 GTGGCAATCATTTTGCATGCTTTCACTTAACATTAACATTAAGAAATTCATCTAGG 364
Oy      71 AGTATATTTGCACATATGAAGGTTGATTAAGCCTTAATGACCATATTATAGAC 130
Db      365 AGTATTTCTTACAAATGTGAAGATGTGTAAGCCTTAACTGATGCTCAACCTTACTA 424
Oy      131 AACATTAAAGAACCAAGTAATGATCGACGTTAATATGACAGTGCAGTTGTGATA 190
Db      425 AACATTAAGTAATTCATCTGAGAAAAACCTTAATAATGT-----GAAGATGTGGCA 478
Oy      191 AAGCATTTTTCAGAAATGACATTTGGAAAGCATATGTATACATTCGAAAAAAG 250
Db      479 AAGCTTTAACCGGTCTCAATCTTACTTAACATTAATAATTAATTCATCTGAGAAAC 538

```

QY 251 CATTCATTGTTCAGTGTGTGTAAGGGGTTAATCTCGACAACTTGAAAAAGACATG 310  
 DB 539 CCTACAAATGTGAAGATGTGGCAAGCTTTAAACGGTCTCAACCTTACTTAACATA 598  
 QY 311 AATTCACCCATCAAAAGTCAATTTAAATGTACATTTGAAAATTTGCAAGAACATTTATA 370  
 DB 599 AAGAAATCTTACAGAAAGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTAAAC 658  
 QY 371 AACATCAATCTT---AAGACATCATATATATCTGTTCATGAAAAAACAATTAAACGTGA 427  
 DB 659 AGTTCCTGATTTCTTAATTAACATTAAGGATTCATATGAAAGTAAACCTTACAAATGTG 718  
 QY 428 AACATGTATAAAGTTTCTGACCTGACCTTCAAAATTAGCAACAATTAATTAATTAATC 487  
 DB 719 AAGAAATGTGCAAAAGCTTTAGAGTATCTCAATTTCTTAATAAACAATAGATATCATTA 778  
 QY 488 ATGTGATCTCC 500  
 DB 779 CTGGGAAAAAAC 791

RESULT 15  
 ADB63624  
 ID ADB63624 standard, cDNA, 2110 BP.  
 XX  
 AC ADB63624;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human cDNA encoding clone THYM20071120.  
 XX  
 KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
 KW tissue regeneration; cell regeneration; membrane protein;  
 KW signal transduction-related protein; transcription-related protein;  
 KW osteoporosis; neurological disease; cancer; tumour.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 170..1792  
 FT /cag= a  
 FT /product= "Clone THYM20071120 protein"  
 PN BP1308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Negai K, Irie R, Tamechika I,  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;  
 XX  
 DR WPI; 2003-450961/43.  
 DR P-PSDB; ADB65594.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX  
 PS Claim 1; Page; 222pp; English.  
 XX  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or

CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
 CC data for this patent is not represented in the printed specification, but  
 CC is based on sequence information supplied by the European Patent Office.  
 XX  
 SQ Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 10; Length 2110;  
 Best Local Similarity 51.7%; Pred. No. 1.5e-08;  
 Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

QY 61 CGTCCCAAAAGATATTTGACATATGAAGGGTGTGATTAAGCTTATATGACCATCA 120  
 DB 1253 CATACCGAAGAAACCTTACATATTAAGATGTGGCAAGCTTTAAACCTCTCA 1312  
 QY 121 TTATTAAGCAACATTAAAGAACCCACAGTATGATGACCGTATTAATGTACGTGAC 180  
 DB 1313 GCCCTTACTACATATGAAGAAATTCACCTGAGAGAAACCCACAAATGT-----GAA 1366  
 QY 181 GATTGTGATTAAGCAATTTTCAAGAAATCAATTGGAACACATATTTGATACATTC 240  
 DB 1367 GAATGTGGCAAGCTTTTAAACCATCTCAAACTTGAACATTAAGAAACCTTCATACT 1426  
 QY 241 GAAAAAACAACATTCATGTTTCAAGTGTGTAAGGGGTTAATCTGCAACAACCTTG 300  
 DB 1427 GGAAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTATCCAAATCTCAAACTT 1486  
 QY 301 AAAAGCATGAATCAACCATCAAAAGTCAATTTAAATGTACATTTGAAAATTTGCAAGAA 360  
 DB 1487 ACTGAACATTAATAAATTTCTATCTGAGAGATACCTTACAGTGTGAAGATGTGGCAA 1546  
 QY 361 GCATTTTATTAACATCAATCTTTA---AAGCATCATATTAATCTGTCTATGAAAAACA 417  
 DB 1547 GCTTTTAAACACTCTCATCTCTTACTACACATTAAGAAATTCATCTGGGAGAAACCC 1606  
 QY 418 TTAACTGTAAACAATGTATTAAGTTTCACTCGACCTTCAAAATTTAGCAACAACATTA 477  
 DB 1607 TACAAATGTGAAGATGTGGCAAGCTTTTAAAGCTTCAAAACCTTACTGAACATTAAG 1666  
 QY 478 TTAAACATCATGATG-GATCTCTGCTTATCATATGATGATCATCTGTGTTTAAAAA 536  
 DB 1667 ATATTTCAATCTGAGAGAAACCTTAAATGTGAAGATGTGCAAAAGCTTTTAAACAA 1726  
 QY 537 TTTCCAACTGTGTGATTTTACATTTTCAATTAATAAACAATGATCAAAACCTTAAGT 596  
 DB 1727 TCTGCAAACT-----TACTAAACATTAATAAATCATATCTGGAAGAACTACAGAAC 1780  
 QY 597 TCTTAATGTGTAAGGTGTGTTGGAAAAAGTTTATCTTCAATATGTTAAGTCA 656  
 DB 1781 TGAATGTGTGATTAATGATTTTGACAACACTTCAAAATTTTCTTAATTAAGGAATCA 1840  
 QY 657 TGATGAT 663  
 DB 1841 TACTGCT 1847

Search completed: February 9, 2005, 02:39:49  
 Job time : 647.722 secs



**This Page Blank (uspto)**



QY 976 GCTTCTTGGAGGTTCAAGAGTGTCTTCTAACTTATCTGATATGTGGAGGAAGATC 1035  
DB 364 GCTTCTTGGAGGTTCAAGAGTGTCTTCTAACTTATCTGATATGTGGAGGAAGATC 423  
QY 1036 AATTGCTTGAAGATAATGTGTAGTATGTTTCTAGAGAAATATGATTTAGTGCACAT 1095  
DB 424 AATTGCTTGAAGATAATGTGTAGTATGTTTCTAGAGAAATATGATTTAGTGCACAT 483  
QY 1096 TTGAAATGCGATGATGATTAATTACAAAGATTGACTCATCTTTAAATAGTATGAAAA 1155  
DB 484 TTGAAATGCGATGATGATTAATTACAAAGATTGACTCATCTTTAAATAGTATGAAAA 543  
QY 1156 GAAAGAACTCCAGAGGTGAACCATTTGGTTAAAAAAGCCAGATGATTTATGCCAAT 1215  
DB 544 GAAAGAACTCCAGAGGTGAACCATTTGGTTAAAAAAGCCAGATGATTTATGCCAAT 603  
QY 1216 GAAACATGAGTATTTCTCGA 1236  
DB 604 GAAACATGAGTATTTCTCGA 624

## RESULT 2

US-09-248-796A-5186  
Sequence 5186, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 5186  
LENGTH: 462  
TYPE: DNA  
ORGANISM: Candida albicans  
US-09-248-796A-5186

Query Match 31.6%; Score 391; DB 4; Length 462;  
Best Local Similarity 98.5%; Pred. No. 4.1e-80;  
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCA 60  
DB 52 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCA 111  
QY 61 CGTCCCAAAAAGTATATTTGACATATGAAGGGTGTATGAAGCCATATATGACCATCA 120  
DB 112 CGTCCCAAAAAGTATATTTGACATATGAAGGGTGTATGAAGCCATATATGACCATCA 171  
QY 121 TTATTAAGCAACATTTAAGAACCCAGTAATGATGACCGTATTAATGTACAGTGAC 180  
DB 172 TTATTAAGCAACATTTAAGAACCCAGTAATGATGACCGTATTAATGTACAGTGAC 231  
QY 181 GATTGTGATTAAGCATTTTTCAGAAAATCACTTTGAAAACACATTTGTATCACATTCC 240  
DB 232 GATTGTGATTAAGCATTTTTCAGAAAATCACTTTGAAAACACATTTGTATCACATTCC 291  
QY 241 GAAAAAACAATTCATTTGAGTGTGTGTGAAGGGGTTAATTCTGACAACACTTG 300  
DB 292 GAAAAAACAATTCATTTGAGTGTGTGTGAAGGGGTTAATTCTGACAACACTTG 351  
QY 301 AAAAGCATGAATCAACCATACAAAGTCAATTTAAATGTACATTTGAAAATGTCAAGA 360  
DB 352 AAAAGCATGAATCAACCATACAAAGTCAATTTAAATGTACATTTGAAAATGTCAAGA 411  
QY 361 GCATTTTAT-AAAACATCAATCTTTAAGACATCATATATATCTGTCATGA 410

DB 412 ACATTTTATAAACAATCAATCTTTAAGACATCATATATATCTGTCATGA 462

## RESULT 3

US-09-016-434-1054  
Sequence 1054, Application US/09016434  
Patent No. 6500938  
GENERAL INFORMATION:  
APPLICANT: Janice Au-Young  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1490  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,434  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0002 US  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1054:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2320 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: g1017721  
US-09-016-434-1054

Query Match 7.8%; Score 96.6; DB 4; Length 2320;  
Best Local Similarity 51.3%; Pred. No. 1.4e-12;  
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGATATATTTGACATATGAAGGGTGTATGAAGCCATATATGACCATCATATTTAGA 128  
DB 964 AAGAAACCCATCAATGTAAAGAAATGTGTAAAGCTTTTAAACGATTTCAACCTTAC 1023  
QY 129 GCAACATTTAAGAACCCAGTAATGATGACCGTATTAATGTACAGTGACGATTTGCA 188  
DB 1024 TACCATGAAAAAATTCATGAGAGAAACCTTACAAATGT-----GAAGAAATGTG 1077  
QY 189 TAAAGCATTTTCAAAAAATCAATTTGGAACACATATTTGATCAATTCGAAAAAAA 248  
DB 1078 CAAGCGTTTAAAGCAGTCTCAACCTTACATCAATGAATATTCATGAGAGAGA 1137  
QY 249 ACCATTGATTTGTCAGTGTGTGTAAAGGGGTTAATCTCGACAACACTTGAAAAAGA 308  
DB 1138 ACCCTAACAATGTAAAGAAATGTGAAAGCCCTTTAAACAGTGTGACACCTTACACA 1197  
QY 309 TGAATTCACCATACAAAGTCAATTTAAATGTACATTTGAAAAATGTCAAGAAAGCATTTTA 368

Db 1198 TGAAGTAATTCATACGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCTTTAA 1257  
Qy 369 TAACA---TCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAAATTAACGCTG 425  
Db 1258 TCATTTCTCAGACCTTACTACATAGATTAATTCATCTGAGAGAAACCTTACAAATG 1317  
Qy 426 TAAACAATGTAAATTAAGTTTTCCTGACCTTCAAAAATTAGCAACAATTAATTAACA 485  
Db 1318 TAAAGAAATGTGTAAGCTTTTAAACATCTTCAACCTTACATAAATAGATTAATCA 1377  
Qy 486 TCATGTGATCTCTGCTTATCATATGATCATCTGCTGTTGTTTAAAAATTTCCAAAC 545  
Db 1378 TACTGAGAGAGAACCTTACAAATCTAAAGATGTGAAAAAGCTTTTAAACCAATCTCAAA 1437  
Qy 546 TTGTCAGATTAATTCATATTAATAAACAAGTCATCCAAAATTAATGCTCTAATG 605  
Db 1438 ACTTACTGAACATTAAGAAATTCATATCTGAGAGAAACCTTA---TGAATGTAAAAATG 1494  
Qy 606 TGGTAAAGTGTGTGTTGGGAAAAAGTTTATCTTCATATGTTAAAGTCATGATGA 662  
Db 1495 TGGCAAAAGCTTTTAAACCAAGTCCCAATCTTACTAGACATTAAGAAAGTCATACAGA 1551

## RESULT 4

US-09-949-016-4204  
; Sequence 4204, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4204  
; LENGTH: 3798  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-4204

Query Match 7.3%; Score 90.8; DB 4; Length 3798;

Best Local Similarity 53.9%; Pred. No. 3.4e-11; Indels 9; Gaps 2;  
Matches 235; Conservative 0; Mismatches 192;

Qy 69 AAGATATATTTGACATATGAGGGTGTGATTAAGCCTATATGACCATCATTTATTAGA 128  
Db 1848 AGAGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTTAGCATTCTTCAACCTTGC 1307  
Qy 129 GCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGTGCAGCATTTGTA 188  
Db 1908 TAAACATTAAGAGATTCATCTGAGAGAAACCTTCAAAATGT-----GAAAGATGTGG 1961  
Qy 189 TAAAGATTTTGAAGAAATCAATTTGAAACACATATGTATCATTCGGAATAAAAA 248  
Db 1962 CAAGGCTTTAGCGCTTCTTCAACCTTGTAAACATTAAGAAATTCATATCTGAGAGAA 2021  
Qy 249 ACCATTTCATTTGTCAGTGTGTGTAAGGGGTTAAATTCGCAACAACCTTGAAGAGACA 308  
Db 2022 ACCCTCAAAATGTAAAGATGTGGCAAAAGCTTTTGAATTCCTCAACCTTGTATATCA 2081  
Qy 309 TGAATATCACCATTAACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTTTA 368  
Db 2082 TAAAGATTAATCAATCTGAAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTAA 2141  
Qy 369 TAAACATCATC---TTTAAGACATCATATATATATCTGTTCAAGAAAAACATTTAACGCTG 425

Db 2142 GCGACTCTCAACCTTACTTAACATTAATAATTAACATGCTGAGAGAAACCTTCAAAATG 2201  
Qy 426 TAAACATGTAAATTAAGTTTTCCTGACCTTCAAAAATTAGCAACAATTAATTAACA 485  
Db 2202 TGAAGAAATGTGGCAAAAGCTTTTAAATGATCTTCAAAATCTTACTATATTAAGTTTATCA 2261  
Qy 486 TCATGTGATCTCTCT 501  
Db 2262 TACTGAGAGAAACCT 2277

## RESULT 5

US-09-949-016-12227  
; Sequence 12227, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12227  
; LENGTH: 156942  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(156942)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12227

Query Match 7.3%; Score 90.8; DB 4; Length 156942;

Best Local Similarity 53.9%; Pred. No. 1.1e-10; Indels 9; Gaps 2;  
Matches 235; Conservative 0; Mismatches 192;

Qy 69 AAGATATATTTGACATATGAGGGTGTGATTAAGCCTATATGACCATCATTTATTAGA 128  
Db 152999 AGAGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTTAGCATTCTTCAACCTTGC 153058  
Qy 129 GCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGTGCAGCATTTGTA 188  
Db 153059 TAAACATTAAGAGATTCATCTGAGAGAAACCTTCAAAATGT-----GAAAGATGTGG 153112  
Qy 189 TAAAGATTTTGAAGAAATCAATTTGAAACACATATGTATCATTCGGAATAAAAA 248  
Db 153113 CAAGGCTTTAGCGCTTCTTCAACCTTGTAAACATTAAGAAATTCATATCTGAGAGAA 153172  
Qy 249 ACCATTTCATTTGTCAGTGTGTGTAAGGGGTTAAATTCGCAACAACCTTGAAGAGACA 308  
Db 153173 ACCCTCAAAATGTAAAGATGTGCAAAAGCTTTTGAATTCCTCAACCTTGTATATCA 153232  
Qy 309 TGAATATCACCATTAACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTTTA 368  
Db 153233 TAAAGATTAATCAATCTGAAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTAA 153292  
Qy 369 TAAACATCATC---TTTAAGACATCATATATATATCTGTTCAAGAAAAACATTTAACGCTG 425  
Db 153293 GCGACTCTCAACCTTACTTAACATTAATAATTAACATGCTGAGAGAAACCTTCAAAATG 153352  
Qy 426 TAAACATGTAAATTAAGTTTTCCTGACCTTCAAAAATTAGCAACAATTAATTAACA 485  
Db 153353 TGAAGATGTGGCAAAAGCTTTTAAATGATCTTCAAAATCTTACTATATTAAGTTTATCA 153412

QY 486 TCATGTGATCTCCT 501  
DB 153413 TACTGGAGAGAAACCT 153428

## RESULT 6

US-09-949-016-15946  
Sequence 15946, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15946  
LENGTH: 156950  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(156950)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15946

Query Match 7.3%; Score 90.8; DB 4; Length 156950;  
Best Local Similarity 53.9%; Pred. No. 1.1e-10;

Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAAGGGTGTATTAAGCCTTAATCGACCATTTATTAGA 128  
DB 152999 AGAGAAACCTTCAATGTGAGAAATGTGGCAAGCTTTTACCATCTTCAACCTTGC 153058  
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGACGATTGTGA 188  
DB 153059 TAAACATAGAGAAATCATCTGAGAGAAACCTTCAATGT-----GAAGAAATGTGG 153112  
QY 189 TAAAGATTTTCAAGAAATTCATTTGGAACACATATTGTATCATCTCCGAAAAAAA 248  
DB 153113 CAAAGCTTTTACCGCTTCTTCAACCTTCTTAAACATTAAGAAATTCATCTGAGAGAA 153172  
QY 249 ACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAATCTCGACAAACATTTGAAAAAGA 308  
DB 153173 ACCCTTAATATGTAAAGATGTGGCAAGCTTTTACCAATCTCTCAACCTTGTATATCA 153232  
QY 309 TGAATTCACCCATACAAAGTCATTTAATGTACATTTGAAAAATGTCAAGAAAGCATTTTA 368  
DB 153233 TAAAGATTACTACTGAGAGAAACCTTACAAATGTAAGATGTGACAAACCTTTTA 153292  
QY 369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCAAGAAAAACATTAACGTG 425  
DB 153293 GCGACTCTCAACCTTACTTAACTTAATATACATGTCTGAGAGAAACCTTACAAATG 153352  
QY 426 TAAACATGTAATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAATAACA 485  
DB 153353 TAAAGATGTGGCAAGCTTTTAATGATCTTCAAAATCTTACTATACATTAAGTTTATCA 153412  
QY 486 TCATGTGATCTCCT 501  
DB 153413 TACTGGAGAGAAACCT 153428

RESULT 7  
US-09-949-016-485

Sequence 485, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 485  
LENGTH: 3839  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-485

Query Match 7.2%; Score 89.2; DB 4; Length 3839;  
Best Local Similarity 53.7%; Pred. No. 7.9e-11;  
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAAGGGTGTATTAAGCCTTAATCGACCATTTATTAGA 128  
DB 964 AGAGAAACCTTCAATGTGAGAAATGTGGCAAGCTTTTACCATTTTCAACCTTGC 1023  
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGACGATTGTGA 188  
DB 1024 TAAACATAGAGAAATCATCTGAGAGAAACCTTCAATGT-----GAAGAAATGTGG 1077  
QY 189 TAAAGATTTTCAAGAAATTCATTTGGAACACATATTGTATCATCTCCGAAAAAAA 248  
DB 1078 CAAAGCTTTTACCGCTTCTTCAACCTTCTTAAACATTAAGAAATTCATCTGAGAGAA 1137  
QY 249 ACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAATCTCGACAAACATTTGAAAAAGA 308  
DB 1138 ACCCTTAATATGTAAAGATGTGGCAAGCTTTTACCAATCTCTCAACCTTGTATATCA 1197  
QY 309 TGAATTCACCCATACAAAGTCATTTAATGTACATTTGAAAAATGTCAAGAAAGCATTTTA 368  
DB 1198 TAAAGATTACTACTGAGAGAAACCTTACAAATGTAAGATGTGACAAACCTTTTA 1257  
QY 369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCAAGAAAAACATTAACGTG 425  
DB 1258 GCGACTCTCAACCTTACTTAACTTAATATACATGTCTGAGAGAAACCTTACAAATG 1317  
QY 426 TAAACATGTAATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAATAACA 485  
DB 1318 TGAAGATGTGGCAAGCTTTTAATGATCTTCAAAATCTTACTATACATTAAGTTTATCA 1377  
QY 486 TCATGTGATCTCCT 501  
DB 1378 TACTGGAGAGAAACCT 1393

RESULT 8  
US-09-949-016-27359/c  
Sequence 27359, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20

;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 27359  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-27359

Query Match  
Best Local Similarity 51.4%; Pred. No. 1.6e-10;  
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

QY 7 GAAAGTGCAGAAACCAATGATATCATCTTTAATATCTTCTTCTTCATGACGCTCC 66  
DB 538 GAATGTGGCAAGCTTTTAAGCACTCCTCAGCCCTTCTAAGCATTAATAATATCATGCT 479  
QY 67 AAAAGTATATTTGACATATGAAAGGTGATTAAGCCTAATATGACCATCATTTATTA 126  
DB 478 GGAAGAACTCTACAAATGTGAGAAATGTGCAAAAGCTTTATCATCTTCAATCTT 419  
QY 127 GAGCAACATTTAAGAACCCACAGTAATATGACCCGTATTAATGTACAGTGAAGATTTG 186  
DB 418 ACACACATTAAGTATATTTATCTTAAGAGAAACCTTCCAGAGT-----GAAAGATGT 365  
QY 187 GATTAAGCATTTTTCAGAAAAATCATTTTGAAAAACATATTTGTATCATATTCGAAAA 246  
DB 364 GACAAACATTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAGAG 305  
QY 247 AAACCATTCATTTGTTCAAGTGTGTGTAAGGGGTTAATCTCGACAAACCTTGAAAA 306  
DB 304 AAAMCCTACAAATGTGAAAGATGTGGCAAGCATTTTGAAGCCCTTCAACCTTACTACA 245  
QY 307 CATGAATCAACCCATACAAAGTCAATTAATGTACATTTGAAATTTGCAAGAAAGCATTT 366  
DB 244 CATAGAGATGACACATCGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTT 185  
QY 367 TATTAACATCAATC---TTTAAGACATCATATATATCTGTTCATGAAAAAATTAACG 423  
DB 184 AGCAATCTCTCAACCTTACTACACATTAATATTTACTGAGAGAAACCTTACAAA 125  
QY 424 TGTAAACATGTAAATTAAGTTTCACTCGACCTTCAAAATTAAGCAACATTAATTA 483  
DB 124 TGTGAAGAAATGTGGCAAGCTTTTAAGAAATCTTCACTTACTGAACATTAAGATTA 65  
QY 484 CATCATGTG 493  
DB 64 CATACTGGAG 55

RESULT 9  
US-09-949-016-149638/c  
;; Sequence 149638, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: C1001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; PRIOR FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 149638

;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-149638

Query Match  
Best Local Similarity 51.4%; Pred. No. 1.6e-10;  
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

QY 7 GAAAGTGCAGAAACCAATGATATCATCTTTAATATCTTCTTCTTCATGACGCTCC 66  
DB 538 GAATGTGGCAAGCTTTTAAGCACTCCTCAGCCCTTCTAAGCATTAATAATATCATGCT 479  
QY 67 AAAAGTATATTTGACATATGAAAGGTGATTAAGCCTAATATGACCATCATTTATTA 126  
DB 478 GGAAGAACTCTACAAATGTGAGAAATGTGCAAAAGCTTTTATCATCTTCAATCTT 419  
QY 127 GAGCAACATTTAAGAACCCACAGTAATATGACCCGTATTAATGTACAGTGAAGATTTG 186  
DB 418 ACACACATTAAGTATATTTATCTTAAGAGAAACCTTCCAGAGT-----GAAAGATGT 365  
QY 187 GATTAAGCATTTTTCAGAAAAATCATTTTGAAAAACATATTTGTATCATATTCGAAAA 246  
DB 364 GACAAACATTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAGAG 305  
QY 247 AAACCATTCATTTGTTCAAGTGTGTGTAAGGGGTTAATCTCGACAAACCTTGAAAA 306  
DB 304 AAAMCCTACAAATGTGAAAGATGTGGCAAGCATTTTGAAGCCCTTCAACCTTACTACA 245  
QY 307 CATGAATCAACCCATACAAAGTCAATTAATGTACATTTGAAATTTGCAAGAAAGCATTT 366  
DB 244 CATAGAGATGACACATCGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTT 185  
QY 367 TATTAACATCAATC---TTTAAGACATCATATATCTGTTCATGAAAAAATTAACG 423  
DB 184 AGCAATCTCTCAACCTTACTACACATTAATATTTACTGAGAGAAACCTTACAAA 125  
QY 424 TGTAAACATGTAAATTAAGTTTCACTCGACCTTCAAAATTAAGCAACATTAATTA 483  
DB 124 TGTGAAGAAATGTGGCAAGCTTTTAAGAAATCTTCACTTACTGAACATTAAGATTA 65  
QY 484 CATCATGTG 493  
DB 64 CATACTGGAG 55

RESULT 10  
US-09-949-016-44799  
;; Sequence 44799, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: C1001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; PRIOR FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 44799  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-44799

Query Match  
Best Local Similarity 54.4%; Pred. No. 2.1e-10;

Matches 241; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 61 CGTCCAAAAGATATTTGACATATGAAGGTGTGATGAAGCCATATATGACCATCA 120  
DB 56 CATACTAGAGAAACCCCTACAAATGTGAAGATGTGGAGGCATTTATATGCTCTCA 115  
QY 121 TTATTAAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAC 180  
DB 116 ACCCTAATGACATATGAAGAGATGACACTGGAGAGAAACCCCTACAAATGT-----GAA 169  
QY 181 GATTGATGAAGCATTTTTCAGAAAATACATTTGAAAACATTTGTATCATCATTC 240  
DB 170 GAATGTGCAAAATCTTTAGCCATCTCAACCCCTTACTACATATGAATTAATCTACT 229  
QY 241 GAAAAAACAATTCATTTGTCAGTGTGTGTGAAGGGTTAATCTGCACAACTTG 300  
DB 230 GGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCTTTAATGCTGTCACTCT 289  
QY 301 AAAAGACATGAATTCACCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAA 360  
DB 290 ACTAAACATTAATATTAATCTACTGAAGAGAAACCCCTACAAATGTGAATAATGTGGCAAA 349  
QY 361 GCATTTTA-TAAACATCATCTTTA--AGACATCATATATTAATCTGTGATGAATAACA 417  
DB 350 GCCTTTAAGCAGCTTCAATCTTACTAACCATAAGAGAAATTCATGAGAGAAACCC 409  
QY 418 TTACGCTGAACAAATGATTAAGTTTCACTGACCTTCAAAATTAACACAACTATAA 477  
DB 410 TATTAATGTGAAGATGTGGCAAAATCTTTTAACCGGTCTTCACTTTACTAAACATTAAG 469  
QY 478 TTAACATCATGTGTGATCTCC 500  
DB 470 GTAATTCATACCTGAGATAAACCC 492

RESULT 11  
US-09-949-016-13027  
; Sequence 13027, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13027  
; LENGTH: 45138  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(45138)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13027

Query Match 7.0%; Score 86.2; DB 4; Length 45138;  
Best Local Similarity 54.4%; Pred. No. 8.4e-10;  
Matches 241; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 61 CGTCCAAAAGATATTTGACATATGAAGGTGTGATGAAGCCATATATGACCATCA 120  
DB 42121 CATACTAGAGAAACCCCTACAAATGTGAAGATGTGGAGGCATTTATATGCTCTCA 42180  
QY 121 TTATTAAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAC 180

DB 42181 ACCCTAATGACATTAAGAGGATGCACTGGAGAGAAACCCCTACAAATGT-----GAA 42234  
QY 181 GATTGATGAAGCATTTTTCAGAAAATCACTTTGAAAACATATGTATCATCATTC 240  
DB 42235 GAATGTGCAAAATCTTTAGCCATCTCAACCCCTTACTACATTAAGATTAATCTACT 42294  
QY 241 GAAAAAACAATTCATTTGTCAGTGTGTGTGAAGGGTTAATCTGCACAACTTG 300  
DB 42295 GGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCTTTAATGCTGTCACTCT 42354  
QY 301 AAAAGACATGAATTCACCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAA 360  
DB 42355 ACTAAACATTAATATTAATCTACTGAAGAGAAACCCCTACAAATGTGAATAATGTGGCAAA 42414  
QY 361 GCATTTTA-TAAACATCATCTTTA--AGACATCATATATTAATCTGTGATGAATAACA 417  
DB 42415 GCCTTTAAGCAGCTTCAATCTTACTAACCATAAGAGAAATTCATGAGAGAAACCC 42474  
QY 418 TTACGCTGAACAAATGATTAAGTTTCACTGACCTTCAAAATTAAGCACAATATAA 477  
DB 42475 TATTAATGTGAAGATGTGGCAAAATCTTTTAACCGGTCTTCACTTTACTAAACATTAAG 42534  
QY 478 TTAACATCATGTGTGATCTCC 500  
DB 42535 GTAATTCATACCTGAGATAAACCC 42557

RESULT 12  
US-09-949-016-27360/c  
; Sequence 27360, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 27360  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-27360

Query Match 6.9%; Score 85.6; DB 4; Length 601;  
Best Local Similarity 51.4%; Pred. No. 2.9e-10;  
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATTCATATCATCTTTAATATCTTCTTTCATCAGCTCC 66  
DB 499 GAATGTGCAAAAGCTTTTAAGCATCTCTGAGCCCTTGCAAAATTAATTAATCATGT 440  
QY 67 AAAAGATATTTGACATATGAAGGTGTGATGAAGCCATATATGACCATATTAATTA 126  
DB 439 GGAGAGAAACCTACAAATGTGAAGATGTGGCAAGCCTTTAATCAATCTTCAAAATCT 380  
QY 127 GAGCAATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAGCATTTGT 186  
DB 379 ACGACATATTAAGATTAATTCATCTAAGAGAAACCTTCCAGAGT-----GAAAGATGT 326  
QY 187 GATTAAGATTTTTCAGAAAATCACTTTGAAAACATATTTGTATCATCTTCCGAAAAA 246  
DB 335 GACAAAGCATTTATCTGTGCTCTCACCCTTACTGAACAATGAAGATTCATACAGAGAG 266



QY 247 AAACATTCATGTTGAGTGTGTGTAAAGGGGTTAATCTCGACAAACCTTGAAAAGA 306  
DB 265 AAAACCTTAAATGTGAAGATGTGGCAAGCATTTAGCCAGCTTGACACCTTACTACA 206  
QY 307 CATGAATTCACCCATACAAAGTCAATTTAATGTACATTTGAAAATTGTCAAGAACATTT 366  
DB 205 CATTAAGGATGTGCACTGAGAGAAACCTTACAAAGTGAAGATGTGGCAAGCTTTT 146  
QY 367 TATAACATCATC---TTTAAGACATCATATATATCTGTTCATGAAAACATTAACG 423  
DB 145 AGCCAACTCTCAACCTTACTACACATTAAGATTAATCTACTGAGAGAAACCTTACAA 86  
QY 424 TGTAAACATGTATAAAGTTTTCATCTGACCTTCAAAATTAAGCAACAATAATTAAA 483  
DB 85 TGTGAAGATGTGGCAAGCTTTTGAAGAAATCTTCACTCTTACTGAACATTAAGATTA 26  
QY 484 CATCATGTGTG 493  
DB 25 CATACTGGAG 16

## RESULT 13

US-09-949-016-149637/c  
Sequence 149637, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 149637  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-149637

Query Match 6.9%; Score 85.6; DB 4; Length 601;

Best Local Similarity 51.4%; Pred. No. 2.9e-10;  
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAGTGAAGCAACCAATGATATCATCTTTAATATCTTCTTCTTATCATGCGTCC 66  
DB 499 GAATGTGGCAAACTTTTAAGCACTCTCGACCCCTTGCTAAACATAAATATATCATGCT 440  
QY 67 AAAAGTATATTGTCATATGAAGGGTGTGATTAAGCTATATGACCATCATTTATTA 126  
DB 439 GGAAGAAATCTTACAAATGTGAGATGTGGCAAGCTTTTATCATCTTCAAACTTT 380  
QY 127 GAGCAACATTTAAGAACCCACAGTATGATGACCCGTATTAATGTACATGAGCATTTG 186  
DB 379 AGCACACATTAAGATTAATTTACTTAAGAGAAACCTTCCAAAGT-----GAAAGATGT 326  
QY 187 GATTAAGCATTTTTCAGAAAATCACATTTGAAAACATATTTGTATACATTTCCGAAAA 246  
DB 325 GACAAAGCATTTATCTGTCTCTCARCCCTTACTGAACATTAAGAAATTCATACAGAGAG 266  
QY 247 AAACATTTCCATGTTGAGTGTGTGTAAAGGGGTTAATCTCGACAAACCTTGAAAAGA 306  
DB 265 AAAACCTTAAATGTGAAGATGTGGCAAGCATTTAGCCAGCTTGACACCTTACTACA 206  
QY 307 CATGAATTCACCCATACAAAGTCAATTTAATGTACATTTGAAAATTGTCAAGAACATTT 366  
DB 205 CATTAAGGATGTGCACTGAGAGAAACCTTACAAAGTGAAGATGTGGCAAGCTTTT 146

QY 367 TATAACATCATC---TTTAAGACATCATATATATCTGTTCATGAAAACATTAACG 423  
DB 145 AGCCAACTCTCAACCTTACTACACATTAAGATTAATCTACTGAGAGAAACCTTACAA 86  
QY 424 TGTAAACATGTATAAAGTTTTCATCTGACCTTCAAAATTAAGCAACAATAATTAAA 483  
DB 85 TGTGAAGATGTGGCAAGCTTTTGAAGAAATCTTCACTCTTACTGAACATTAAGATTA 26  
QY 484 CATCATGTGTG 493  
DB 25 CATACTGGAG 16

## RESULT 14

US-09-949-016-27361/c  
Sequence 27361, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 27361  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-27361

Query Match 6.7%; Score 83.2; DB 4; Length 601;

Best Local Similarity 52.5%; Pred. No. 1e-09;  
Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

QY 57 ATCACTGCCCAAAAGTATATTTGCAATATGAAGGGTGTGATTAAGCCCTTAATGAC 116  
DB 554 AACTCATCTGAAGAGAAACCTTACAAATGTAAAGATGTGACAAACCTTTAAGCGACT 495  
QY 117 ATCATTTATAGACCAATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGT 176  
DB 494 CTCACCCCTTACTTAACATTAATTAATCATGTCTGAGAGAAACCTTACAAATGT----- 440  
QY 177 GAGCATGTGTATTAAGCATTTTTCAGAAAATCAATTTGAAAACATATTTGATAC 236  
DB 439 -GAAGATGTGGCAAGCTTTTATGATCTTCAAACTTACTATATCATTAAGTTATTTCA 381  
QY 237 TTCGAAAAAAACCATTCATGTGTCAAGTGTGTGTAAGGGGTTAATCTGACACAA 296  
DB 380 TACTGAGAGAAACCTTACAAAGTGTGAAGATGTGGCAAGCTTTAATGCGTCTCAAG 321  
QY 297 CTGAAAAGACATGAATATCCCATCAAAAGTCAATTTAATGTACATTTGAAAATTGTCA 356  
DB 320 CCTTACTTAACATTAAGAAATTTATCTAGAGAGAAACCTTCAAAATGTAAAGATGTGG 261  
QY 357 AGAAGCATTTTATTAACATCAATCTTTA--AGACATCATATATTAATCTGTGATGAAA 413  
DB 260 CAAGCATTTATATGTGCTTCAACCTTAATGACATTAAGAGATTAACACTGAGAGAA 201  
QY 414 AACATTAAGTGTAAACATGTATTAATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACA 473  
DB 200 GCCCTTACAAATGTGAAGATGTGGCAAGCTTTTACCGTCTTCAACCTTACTAAGCA 141  
QY 474 TAAATTAACATCATGTGTG 493

Db 140 TAAAGCAATTCATCTGAG 121

## RESULT 15

US-09-949-016-149636/c  
Sequence 149636, Application US/09949016  
Patent No. 6812339

## GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 149636  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-149636

Query Match 6.7%; Score 83.2; DB 4; Length 601;  
Best Local Similarity 52.5%; Pred. No. 1e-09;

Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

QY 57 ATACGCTCCCAAAAGTATATTTGCACATATGAGGGTGTGATTAAGCTATATGACC 116  
DB 554 AACTCTACTGAGAGAAACCTACAAATGTAAAGATGTGACAAACCTTTAAGGACT 495  
QY 117 ATCATTATTAGAGCAACATTTAAGAACCAAGTAATGATCGACCGTATTAATGACGT 176  
DB 494 CTCACCTTACTTAACATTAATAATATATGATGCTGAGAGAAACCTACAAATGT----- 440  
QY 177 GGACGATTGTGATTAAGCATTTTTCGAAATATCATTTGGAAACATATTTGTATCACA 236  
DB 439 -GAAAGATGTGGCAAGCTTTTAATGATCTTCAAACTTACTATACATTAAGTTATTTCA 381  
QY 237 TTCGAAAAAAACCATTCATTTGTCAGTGTGTTAAAGGGTTAATTCTGACACACA 296  
DB 380 TACTGAGAGAAACCTTACAGTGTGAAGATGTGGCAAAAGCATTTAAGTCTCTCAG 321  
QY 297 CTTGAAAGAGCATGAATGACCCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGCA 356  
DB 320 CCTTACTTAACATTAAGAAATTCATCTAGAGAAACCTTCAAAATGTAAGAAATGTGG 261  
QY 357 AGAAGCATTTTATTAACATCATCTTTA--AGACATCATATTTATCTGTTGATGAAA 413  
DB 260 CAAGCATTTATATGTCTTCAACCTTACATTAAGAGATACACATGAGAGAGAA 201  
QY 414 AACATTACGTGTAACATGTATTAAGTTTCACTCGACCTTCAAAATTTAGCACACACA 473  
DB 200 GCCCTTCAATGTGAAGATGTGGCAAGCTTTTACCGTTCTCAACCTTACTAAGCA 141  
QY 474 TAAATTAAACATCATGTGTG 493  
DB 140 TAAAGCAATTCATCTGAG 121

Search completed: February 9, 2005, 07:51:09  
Job time : 227.728 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 9, 2005, 05:36:40 ; Search time 2154.88 Seconds  
(without alignments)  
3301.382 Million cell updates/sec

Title: US-09-831-804-1\_COPY\_720\_1955  
Perfect score: 1236  
Sequence: 1 atgagtggaagtgacgaac.....aaacatcagtattctcga 1236

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4313806 seqs, 287871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.8	99.7	1239	US-10-032-585-6493	Sequence 6493, Ap
2	99.4	8.0	2132	US-10-094-749-726	Sequence 726, App
3	96.6	7.8	2320	US-10-029-386-20763	Sequence 1054, Ap
4	92	7.4	2873	US-10-029-386-20763	Sequence 20763, A
5	92	7.4	2873	US-10-029-386-20763	Sequence 369, App
6	91.8	7.4	2110	US-10-104-047-1778	Sequence 1778, Ap
7	91.8	7.4	3078	US-10-104-047-1778	Sequence 622, App
8	90.4	7.3	2114	US-10-094-749-692	Sequence 692, App
9	90.2	7.3	2230	US-10-094-749-692	Sequence 499, App
10	89.8	7.3	590	US-10-029-386-13298	Sequence 13298, A
11	89.2	7.2	2221	US-10-108-260A-29	Sequence 29, Appl

12	89.2	7.2	3839	US-10-172-118-910	Sequence 910, App
13	89.2	7.2	3839	US-10-342-887-910	Sequence 910, App
14	89	7.2	3502	US-10-098-841-48	Sequence 48, Appl
15	88.6	7.2	1466	US-10-029-386-25074	Sequence 25074, A
16	88.4	7.2	1410	US-10-029-386-20532	Sequence 20532, A
17	88.4	7.2	2509	US-10-381-321-11	Sequence 11, Appl
18	88.4	7.2	4132	US-10-198-846-13133	Sequence 13133, A
19	87.6	7.1	583	US-10-029-386-13093	Sequence 13093, A
20	87.6	7.1	1757	US-10-203-052B-6	Sequence 6, Appl1
21	87.2	7.1	831	US-10-264-049-975	Sequence 975, Appl
22	86.8	7.0	2662	US-10-363-616-66	Sequence 66, Appl
23	86.2	7.0	1422	US-10-029-386-22636	Sequence 22636, A
24	86.2	7.0	1422	US-10-029-386-20322	Sequence 20322, A
25	86.2	7.0	1725	US-10-029-386-20555	Sequence 20555, A
26	86.2	7.0	1727	US-10-029-386-20367	Sequence 20367, A
27	86.2	7.0	2558	US-09-764-864-1664	Sequence 1664, Ap
28	86.2	7.0	2558	US-09-764-864-1665	Sequence 1665, Ap
29	86	7.0	2664	US-10-108-260A-511	Sequence 511, App
30	86	7.0	2690	US-10-723-860-8039	Sequence 8039, Ap
31	86	7.0	3400	US-10-104-047-1037	Sequence 1037, Ap
32	85.4	6.9	2505	US-10-108-260A-808	Sequence 808, App
33	84.8	6.9	2239	US-10-094-749-1084	Sequence 1084, Ap
34	84.4	6.8	1191	US-10-029-386-22992	Sequence 22992, A
35	84.4	6.8	3138	US-10-104-047-83	Sequence 83, Appl
36	84.2	6.8	1578	US-10-029-386-20214	Sequence 20214, A
37	84	6.8	565	US-10-029-386-3970	Sequence 3970, Ap
38	84	6.8	1619	US-10-422-522-46	Sequence 46, Appl
39	84	6.8	3617	US-10-029-386-22868	Sequence 22868, A
40	84	6.8	5820	US-10-363-616-44	Sequence 44, Appl
41	83.6	6.8	596	US-10-029-386-2664	Sequence 6664, Ap
42	83.2	6.7	1101	US-10-029-386-25405	Sequence 25405, A
43	83.2	6.7	2241	US-10-108-260A-1289	Sequence 1289, Ap
44	83.2	6.7	2525	US-10-108-260A-1358	Sequence 1358, Ap
45	83	6.7	500	US-10-029-386-6818	Sequence 6818, Ap

ALIGNMENTS

RESULT 1					
US-10-032-585-6493					
; Sequence 6493, Application US/10032585					
; Publication No. US20030180953A1					
; GENERAL INFORMATION:					
; APPLICANT: Terry, Roemer D.					
; APPLICANT: Bo, Jlang					
; APPLICANT: Charles, Boone					
; APPLICANT: Howard, Bussey					
; TITLE OR INVENTION: Gene Disruption Methodologies for Drug Target Discovery					
; FILE REFERENCE: 10182-005-999					
; CURRENT APPLICATION NUMBER: US/10/032,585					
; CURRENT FILING DATE: 2001-12-20					
; NUMBER OF SEQ ID NOS: 8000					
; SOFTWARE: PatentIn version 3.1					
; SEQ ID NO 6493					
; LENGTH: 1239					
; TYPE: DNA					
; ORGANISM: Candida albicans					
US-10-032-585-6493					
Query Match					
Best Local Similarity 99.8%; Pred. No. 9.9e-227;					
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
QY	1	ATGAGTGAAGTGACGAACCAATCGATATCACTTTAATATCTTCTTCATCA	60		
DB	1	ATGAGTGAAGTGACGAACCAATCGATATCACTTTAATATCTTCTTCATCA	60		
QY	61	CGTCCCAAAAGTATATTTGACATATGAAGGGTGATTAAGCCCTATATGCACATCA	120		
DB	61	CGTCCCAAAAGTATATTTGACATATGAAGGGTGATTAAGCCCTATATGCACATCA	120		
QY	121	TTATTGAGCAACATTAAAGAACCAAGTAATGATCGACCGTATAAATGACAGTGAC	180		

Db 121 TTATTGGACCAATTTAAAGAACCCACAGTAATGACCGGTAAATGTAACAGGAG 180  
QY 181 GATTGATGAAAGCATTTTTCAGAAAATCAATTTGAAAACATATTGATCAATCC 240  
Db 181 GATTGATGAAAGCATTTTTCAGAAAATCAATTTGAAAACATATTGATCAATCC 240  
QY 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATCTCGACAACCTTG 300  
Db 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATCTCGACAACCTTG 300  
QY 301 AAAAGCATGAATCAACCATCAAGTATTTAATGTAACATTTGAAAATGTCAGAA 360  
Db 301 AAAAGCATGAATCAACCATCAAGTATTTAATGTAACATTTGAAAATGTCAGAA 360  
QY 361 GCATTTTAAACATCAATCTTTTAAAGACATCATATTATCTGTCATGAAAACATTA 420  
Db 361 GCATTTTAAACATCAATCTTTTAAAGACATCATATTATCTGTCATGAAAACATTA 420  
QY 421 ACGTGAACCAATGTAAAGTTTCACTCGACCTTCAAAATTTAGCACACATAATTA 480  
Db 421 ACGTGAACCAATGTAAAGTTTCACTCGACCTTCAAAATTTAGCACACATAATTA 480  
QY 481 AAAACATCATGTCGATCTCTGCTTATCATGTGATCATCTGCTGTTTAAAAATTTC 540  
Db 481 AAAACATCATGTCGATCTCTGCTTATCATGTGATCATCTGCTGTTTAAAAATTTC 540  
QY 541 CAAACTGTGTCAGTATTTACAAATTTCAATTAACCACTGATCCAAACCTTAAATGCTC 600  
Db 541 CAAACTGTGTCAGTATTTACAAATTTCAATTAACCACTGATCCAAACCTTAAATGCTC 600  
QY 601 AAATGTGTAAAGTGTGTGGGAAAAAGGTTTACTTCAATATGTTAACTCATGAT 660  
Db 601 AAATGTGTAAAGTGTGTGGGAAAAAGGTTTACTTCAATATGTTAACTCATGAT 660  
QY 661 GATTCTACCATGATCAAAATATGAGATCTTGTGATATGATGTGGGAAATTTGCAAG 720  
Db 661 GATTCTACCATGATCAAAATATGAGATCTTGTGATATGATGTGGGAAATTTGCAAG 720  
QY 721 AAAAATGATTTAGTGAACATTTAATATCTTCATGATGATGATATCCCTGATGATTA 780  
Db 721 AAAAATGATTTAGTGAACATTTAATATCTTCATGATGATGATATCCCTGATGATTA 780  
QY 781 TTAAGGAAATCTGAAGTGAAGAAATTTAGAAACCTATGATCAAGATCGAAATTAAT 840  
Db 781 TTAAGGAAATCTGAAGTGAAGAAATTTAGAAACCTATGATCAAGATCGAAATTAAT 840  
QY 841 AATTGATGATTTAGAAACAGAAATTTAAAGTGAAGAGATGAAGAGTGAAGAA 900  
Db 841 AATTGATGATTTAGAAACAGAAATTTAAAGTGAAGAGATGAAGAGTGAAGAA 900  
QY 901 GATAGCTGATGAAGAAAGAGTGTATGATGACACTCAATGTCAGCTCAAGATCA 960  
Db 901 GATAGCTGATGAAGAAAGAGTGTATGATGACACTCAATGTCAGCTCAAGATCA 960  
QY 961 ATAAATCATTTACTGCTCTTTGGAAGGTTCAAGAGTGTCTTAACTTATCTGAAT 1020  
Db 961 ATAAATCATTTACTGCTCTTTGGAAGGTTCAAGAGTGTCTTAACTTATCTGAAT 1020  
QY 1021 AGTGGAGAAAGATCAATGCTCTTAAGAAATTAATGATGAATGTTTCTAGAGAAAT 1080  
Db 1021 AGTGGAGAAAGATCAATGCTCTTAAGAAATTAATGATGAATGTTTCTAGAGAAAT 1080  
QY 1081 GATTTAAGTCGACATTTGAAAGGTCATGATTAATTTCAAGAAATGATGATCTTA 1140  
Db 1081 GATTTAAGTCGACATTTGAAAGGTCATGATTAATTTCAAGAAATGATGATCTTA 1140  
QY 1141 AATAGATAGAAAAAGAAAGAACTCCAGAAAGTGAACCATGTTTAAAAAGCAGATG 1200  
Db 1141 AATAGATAGAAAAAGAAAGAACTCCAGAAAGTGAACCATGTTTAAAAAGCAGATG 1200  
QY 1201 GATTTATGCCAAATGAACATCATGATTTCTCGA 1236  
Db 1201 GATTTATGCCAAATGAACATCATGATTTCTCGA 1236

Db 1201 GATTTATGCCAAATGAACATCATGATTTCTCGA 1236  
RESULT 2  
US-10-094-749-726  
Sequence 726, Application US/10094749  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHICO  
APPLICANT: YOSHIKAMA, TSUTOMU  
APPLICANT: OTSUKA, MOTORYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
REFERENCE: 08435/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 726  
LENGTH: 2132  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-094-749-726  
Query Match 8.0%; Score 99.4; DB 17; Length 2132;  
Best Local Similarity 50.6%; Pred. No. 4e-09;  
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;  
QY 7 GAAAGTGACGAAACCAATTCATATCATCTTAAATCTTCTTCTTCAACAGTCCC 66  
Db 1384 GATGTGGCAAGCTTTTAAACCAAGTCTCAATCTTACTACATPAAGAAATTCATACT 1443  
QY 67 AAAAGTATATTGGACATATGAAAGGTTGATTAAGCCTATATCGACCATATTATTA 126  
Db 1444 GAGAGAAATCTCTAATATGTAAGAAATGTGGCAAGCTTTTATCGATCTCAAACTT 1503  
QY 127 GAGCAACATTTAAGAACCCACAGTAATGATGACCGGTAAATGTACAGTGAAGATGT 186  
Db 1504 ACTGAACATPAAGAAATTCATCTGAGAGAAACCTACACATGT-----GAAAGATGT 1557  
QY 187 GATTAAGCATTTTTCAGAAATCACAATTTGAAAACATATTGTATCACATTCGAAAAA 246  
Db 1558 GCGAAAGCCTTTTACCACTCCACACCTGTGTACACATPAAGTATTCATCTGAGAG 1617  
QY 247 AAACATTCATGTTTTCAGTGTGTGTAAAGGGGTTAATTCGCAACACTTGAAGAA 306  
Db 1618 AAACCTTCCATGTGAAGAAATGTGTAAAGCTTTAACCAAGTCTCACACCTTACTTGA 1677  
QY 307 CATGAATCACCATACAAAGTCAATTTAATGTACATTTGAAATTTGTCAAGACATTT 366  
Db 1678 CATTAAGAAATTCATACAGAGAGAAACCTTCAATGTGAAGAAATGTGGCAAGCTTTT 1737  
QY 367 TATTAACATCAATCTTTAA--GACATCATATATTATCTGTTCATGAAAAACATTAAG 423  
Db 1738 AACAGTCTCCAAACCTTACTGAGCATTAAGAAATTCATACTGTGAGAAACCTTACAAA 1797

Db 1438 ACTTACTGACACATTAAGAAAATTCATCTGAGAGAAACCTTA--TGAATGTGAAAATG 1494  
 Oy 606 TGGTAAGGTGTGTGTGGGAAAAAGTTTATCTTCACATATGTAAAGTCATGATGA 662  
 Db 1495 TGGCAAGACTTTTAACGACGTCTCAAAATCTTACTAGACATTAAGAAAAGTCATACAGA 1551  
 RESULT 4  
 US-10-029-386-20763/c  
 ; Sequence 20763, Application US/10029386  
 ; Publication No. US20030194704A1  
 GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 20763  
 ; LENGTH: 994  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC010620.3  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
 ; OTHER INFORMATION: NT HIT: 9413650821, EVALUATE 0.00e+00  
 ; OTHER INFORMATION: EST HUMAN HIT: B1093435.1, EVALUATE 0.00e+00  
 ; OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUATE 0.00e+00  
 US-10-029-386-20763  
 Query Match 7.4%; Score 92; DB 16; Length 994;  
 Best Local Similarity 52.0%; Pred. No. 7.7e-08;  
 Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;



```

: GENERAL INFORMATION:
: APPLICANT: HELIX RESEARCH INSTITUTE
: TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
: FILE REFERENCE: H1-A0105
: CURRENT APPLICATION NUMBER: US/10/104,047
: CURRENT FILING DATE: 2002-03-25
: PRIOR APPLICATION NUMBER:
: PRIOR FILING DATE:
: NUMBER OF SEQ ID NOS: 4096
: SOFTWARE: PatencIn Ver. 2.1
: SEQ ID NO 622
: LENGTH: 3078
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-104-047-622

```

Query Match	7.4%;	Score 91.8;	DB 17;	Length 3078;
Best Local Similarity	52.1%;	Pred. No. 1.3e-07;		
Matches 257;	Conservative	0;	Mismatches 227;	Indels 9;
				Gaps 2

QY	11	GTGACGAACCAAAATCGATATCATCTTTAAATCTTTCTCTCTTCATCAACGCTCCAAA	70
Db	650	GTGGCAATTCATTTTGCATGCTTTTCACATTTAACTCAACATAGAAAATTCATTACTAG	709
QY	71	AGTATATTGCAATATGAGGGTGTATAAAGCCTATATTCAGCCATATTATAGAC	130
Db	710	AGTATTTCTTAACAAATGTAGAAATGTGTAAAGCCTTTAACTGTCTCAACCTTACTA	769
QY	131	AACATTTAAGAACCCACAGTATGATCGACGGTATTAATATGTACAGTGGACATGTGATA	190
Db	770	AACATTAAGATTAATTCATCTGAGAAAAACCTTACAAATGT-----GAGATGTGCA	823
QY	191	AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTTCCGAAAAAAC	250
Db	824	AAGCTTTTAAACCGGCTCTCAATCTTACTTAAACATTAATATATTCATCTGAGAGAAAC	883
QY	251	CATTCATATGTTTCAGTGTGTGTAAGGGTTAATTTTCGACAAACATTTGAAGAACATG	310
Db	884	CCTCAAAATGTGAAATATGTGGCAAAAGCTTTTAAACCGGTCTCAACCTTACTTAACATA	943
QY	311	AAATCACCACATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAACGATTTATA	370
Db	944	AAAGATTTCAATACAGAGAGAAACCTTCAAAATGTGAAGAAATGTGGCAAGGCTTTAAC	1002
QY	371	AACATCATCTTT---AAGACATCATATATATCTGTTCATGAAAAAACATTAAGTGA	427
Db	1004	AGTTCTCGATTTTATATAACCTAAGAGATTCATATGGAACATTAACCTTACAATGTG	1062
QY	428	AACATGTATATAAGTTTTCATCTCGACCTTCAAAATATGACACACATTAATTAACATC	487
Db	1064	AAGAAATGTGGCAAAACCTTTAGAGATATCTCAATCTTAAAAAATTAAGATATTCATA	1122
QY	488	ATGTGTGATCTCC	500
Db	1124	CTGGGGAAAAAC	1136

RESULT 8  
US-10-094-749-692  
Sequence 692, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAOBU  
APPLICANT: NAGAI, KEIICHI

```

1  APPLICANT:  IRIE, RYOTARO
2  APPLICANT:  TAMECHIKA, ICHIRO
3  APPLICANT:  SEKI, NAOHITO
4  APPLICANT:  YOSHITAKA, TISTOMU
5  APPLICANT:  OTSUKA, MOTOYUKI
6  APPLICANT:  NAGAHARI, KENJI
7  APPLICANT:  MASUHO, YASUHIKO
8  TITLE OF INVENTION:  NOVEL FULL-LENGTH CDNA
9  FILE REFERENCE:  084335/0160
10 CURRENT APPLICATION NUMBER:  US/10/094,749
11 CURRENT FILING DATE:  2002-03-12
12 PRIOR APPLICATION NUMBER:  60/350,435
13 PRIOR FILING DATE:  2002-01-24
14 PRIOR APPLICATION NUMBER:  JP 2001-328381
15 PRIOR FILING DATE:  2001-09-14
16 NUMBER OF SEQ ID NOS:  3381
17 SOFTWARE:  PatentIn Ver. 2.1.1
18 SEQ ID NO 692
19 LENGTH:  2114
20 TYPE:  DNA
21 ORGANISM:  Homo sapiens
22 US-10-094-749-692

```

Query Match	7.3%	Score 90.4	DB 17	Length 2114
Best Local Similarity	51.8%	Pred. NC. 2.1e-07		
Matches 258	0	Mismatches 231	Indels 9	Gaps 2

Oy	7	GAAAGTACGAAACCAATGCATTCATCTTTAATCTTCTCTTCATCAGTCC	66
Db	1024	GAATGTGCGAAGCCTTCTACCATTTTCAACCTTACTACATAGSTAAATTCATCT	1083
Oy	67	AAAAGTATATTTGCACATATGAAGGGTGTGATGAAGGCTTAATCGACATATTA	126
Db	1084	GGAGAGAAGCCCTTCAATGTGAAGAAATGTGTAAAGCTTTTAACACCTTCAGCCCT	1143
Oy	127	GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTAAATGTAACGTGACGATGT	186
Db	1144	ACTACACATAAGTTCATTCAATGTAAGAAAAAACCTTCAATATG-----GAAGAAATG	1197
Oy	187	GATAAAGCATTTTTCAGAAAATCATTGTGAAACACATATTTGATCATTTCCGAAAA	246
Db	1198	GACAAAGCTTTAAACGATCTCTCATCCCTTACTAAACATTAAGTAATTCATTCTGGAAAG	1257
Oy	247	AAACCATTTCCATGTCAGTGTGTGATGAAGGGGTAAATTCGACAAACCTTGAAAAA	306
Db	1258	AAATCTTCAATGTGAACAATGTGGCAAGGCTTTAATCTGTCTTCAACCTTACAAA	1311
Oy	307	CATGAATTCACCCATCAAAAGTCATTTAAATGTACATTTGAAAAATGTCAAGAACATTT	366
Db	1318	CATGAAGAAATTCATCTGAGAGAAACCCACAAATGTGAAGATGTGGCAAGCCTT	1377
Oy	367	TATTAACAT---CAATCTTTAAGACATCAATATTAATCTGTTCAATGAAAAACATTAACG	423
Db	1378	AATGTGCTTCACACCTTACTACACATPAAGATGATTCATCTGAGAGAAACCTTACAA	1437
Oy	424	TGTAAACAATGTAATAAGTTTCACTGACCTTCAAAATTAAGCAACAATAAATTA	483
Db	1438	TGTGAAGAAATGTGCAAGCCTTTAACAACATCTCCAAACCTTAATTAATCAATAGATAAT	1497
Oy	484	CATCATGTGTGATCTCCCT	501
Db	1498	CATACTGTGAGAAACCT	1515

RESULT 9  
US-10-0094-799-499  
Sequence 499, Application US/10094749  
Publication No. US2003021974A1A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAII, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI



APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHITO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOMYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, NASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 499  
LENGTH: 2230  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-094-749-499

Query Match 7.3%; Score 90.2; DB 17; Length 2230;  
Best Local Similarity 50.4%; Pred. No. 2.4e-07;  
Matches 332; Conservative 0; Mismatches 313; Indels 14; Gaps 4;

QY 7 GAAAGGAGGAAACCAATCGATATCATCTTTAATCTCTCTCTCATACGCTCC 66  
DB 656 GAATGTGACAAATCACTTGGCATGCTTTGACGCTTCAACATTAATAAATTCATCT 715  
QY 67 AAAAATATATTTGACATATGAAAGGTGATTAAGCCTATATGACCATCATATTA 126  
DB 716 AAGAGAAATTTCTACAAATGTGAAGGTGTGAAAACTTTAATCTGTCTCAAACTT 775  
QY 127 GAGCAACATTTAAGAACCAAGTATGATCGACGCTATTAATGTAAGTGAAGATGT 186  
DB 776 TCTAACTTAAGAAATTCATACCTGAGAAAAACCTTAAGATGAGTA-----TGT 829  
QY 187 GATTAAGCATTTTTCAGAAAAATACATTTGAAAAACATATGTATCACTTCCGAAAA 246  
DB 830 GGAAGAGCTTTTCAACAATCTTAACTTAACTTAAGATTAATGCTACTGAGAA 889  
QY 247 AAAACATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGCAACAACCTTGAAGA 306  
DB 890 AAAACCTTAATAATGTGACACCTGTGCAAAACCTTTAAACAGTCTCAACCTTACT 949  
QY 307 CATGAATCAACCCATACAAAGTCATTTAATGTACATTTGAAAAATTGTCAAGAGATT 366  
DB 950 CATTAATATTTATCTAGTGAAGAAACCTTAACAATGTGAACAATGTGGCAAGTCTT 1009  
QY 367 TATTAACATCAATC---TTTAAGACATCATATATATCTGTTATGAAAAAACATTAAG 423  
DB 1010 AAGCAGTCCCAACCCCTTCAACATCAAGTAAATTAATTAATCTGGAAGGAACCATCA 1069  
QY 424 TGTAAACATATGATTAAGTTTCACTCGACCTTCAAAATTTAGCACAACATTAATTA 483  
DB 1070 TGTGAGAAATGTGCAAGCTTTTAACCTATCTTAACTTAATCTGAAACATTAAGAA 1129  
QY 484 CATCATGTG-GATCTCTGCTTATCATATGTGATCATCTGTGTGTTTAAAAATTTCA 542  
DB 1130 TACACTAGAGAGAAAGCTTAACAATGTGAAGATGTGGAAAGCTTTTAACAGTTTCA 1189  
QY 543 AACTGTGTGATTAATCAATTTATTAACAACATGTCATCCAAACTTAATGTCTTA 602  
DB 1190 ACCCTATTACATTAAGATTAATTCATAGCAGAGAAACCC---CACAAATGTGAAGA 1245

QY 603 ATGTGTAAAGGTGTGTTGGAAAAAGGTTTATCTTCATATGTTAAGTCATGATG 661  
DB 1246 ATGTGGAGAGCTTTTAACCACTGCGCAAAAGCTCACTGAACATTAATTCATCTG 1304

## RESULT 10

US-10-029-386-13298  
Sequence 13298, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEWICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Annomax Sequence Listing Engine ver. 1.1  
SEQ ID NO 13298  
LENGTH: 590  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO CHR19.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5  
OTHER INFORMATION: NT HIT: g11475315, EVALU 0.00e+00  
OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALU 1.00e-119  
OTHER INFORMATION: EST\_HUMAN HIT: BG619358.1, EVALU 0.00e+00  
US-10-029-386-13298

Query Match 7.3%; Score 89.8; DB 16; Length 590;  
Best Local Similarity 53.8%; Pred. No. 1.7e-07;  
Matches 224; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

QY 62 GTCCCAAAAAGATATTTGACATATGAAAGGTGATTAAGCCTTAATCGACCATCAT 121  
DB 126 GTGCTTAAGAGAAATCTACAGTGTGAAGATGTGCAAAAGCATTTCTATGTCTTCA 185  
QY 122 TATTAAGACACATTTAAGAACCCACAGTATGATCGACGCTTAATATGATGAGTGC 181  
DB 186 CCTTAATGACATTAAGAGATACACACTGAGAGAAACCTTAAGATGT-----GAAG 239  
QY 182 ATTTGATTAAGCATTTTTCAGAAAAATCATTGGAACACATATGATGACATTTCCG 241  
DB 240 AATGTGGCAAGCTTTTGAAGCTTCTTCAACCTTGTCTAAACATTAAGAGATTCATCTG 299  
QY 242 AAAAACAACATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGCAACAACCTTGA 301  
DB 300 GAGAGAAACCTTCAACAATGTGAAGATGTGCAAAAGCTTTTGAAGCCGTTCTCAACCTTG 359  
QY 302 AAAGACATGAATCAACCCATACAAAGTCATTTAATGTACATTTGAAAAATTGTCAAGAG 361  
DB 360 CTAAACATTAAGAAATTCATCTGAGAGAAACCTTAACAATGTAAAGATGTGGCAAG 419  
QY 362 CATTTTATTAACATCAATTTTAAAGACATCATATATCTGTCTATGAA---AAACAT 418  
DB 420 CTTTAAAGCAATTCCTCAACCCCTTGTATTAAGTAACTATCTAGTGAAGAAACCTCT 479  
QY 419 TAACGTGTAAACATATTAAGTTTCACTCGACCTTCAAAATTTAGCACAACATTAAT 478  
DB 480 ACAAATGTAAAGATGTGCAAAAGCTTTTAAGGACTCTCAACCTTACTTAACAATTA 539  
QY 479 TAAAAATCATGTG 493  
DB 540 TAATCATGCTGAG 554

RESULT 11  
US-10-108-260A-29

```
; Sequence 29, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 2221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-29

Query Match      7.2%; Score 89.2; DB 17; Length 2221;
Best Local Similarity 53.4%; Pred. No. 3,7e-07;
Matches 237; Conservative 0; Mismatches 198; Indels 9; Gaps 2;

QY 61 CGTCCAAAAGATATATTGACATATGAGGGGTGTATAAAGCTTATTCAGACATCA 120
DB 654 CATACCGAGAGAAACCTTTCAGATGTAGAGAAATGGCAAGCTTTATAGAGCTCA 713
QY 121 TTAATTAGAGCAACATTTAAGAACCCACAGTAAATGATGACCGTATTAATGACGTGAC 180
DB 714 TAGCTTACTAGCATCAGAGATGATCTGAGAGAAACCTTTCAAGTGTGTA-----AG 767
QY 181 GATTGTATTAAGCAATTTTTCAGAAATCAATTGGAAACACATATTTGATCATTCC 240
DB 768 GAATGTGGCAAGCTTTTAAACAGAGCTTCAACCTTACTCAACATCAGAAATTCATCT 827
QY 241 GAAAAAAACCATTCCTGTTGCTGCTGTGTGTAAAGGGTTAATTTCACAACTTG 300
DB 828 GAGAGAGAAACCTTCAATGTAAAGATGTGGCAAACTTTTAAACAGGGCTCAGACCTT 887
QY 301 AAAAGACATGAATCAACCCATACAAAGTCAATTTAATGTATGTGAATAATGTCAAGAA 360
DB 888 ACTCGACATCAAGAAATCCATCTGAGAGAGAAAGCTTTCAAATGTGAATGTGGCAAA 947
QY 361 GCATTT---TATAAATCAATCTTTAAGACATCATATATATCTGTTCAAGAAAAACA 417
DB 948 GCTTTTATCTGGGGCTCAGACCTTACTCAACATCAGAGAGTCCATCTAGAGAAATTC 1007
QY 418 TTAACGTGTAAACATGTAATAAAGTTTCACTGACCTTCAAAATTAGCAACATATA 477
DB 1008 TTCAAATGTAAAGATGTGGCAAAAGCTTTTACAGAGAGTCCACCTTACTCAACATCAG 1067
QY 478 TTAAACATCATGTGTGATCTCCT 501
DB 1068 AGAATTCATCTAGAGAGAAACCT 1091

RESULT 12
US-10-172-118-910
; Sequence 910, Application US/10172118
; Publication No. US2003022437A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernades, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
```

```
; SEQ ID NO 910
; LENGTH: 3839
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003430
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-910

Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATTTTGCACATATGAGGGGTGTATAAAGCCTTAATTCAGACCATTAATTAGA 128
DB 964 AGAGAAACCTTACAAATGTGAAGATGTGGCAAAAGCTTTTAGCATTTTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCACAGTAAATGATGACCGTAAATGTAACAGTGCAGATTGTGA 188
DB 1024 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGATGTGG 1077
QY 189 TAAAGCATTTTTCAGAAATTCACATTGGAAACATATTGTATCTCATTCGAAAAAAA 248
DB 1078 CAAAGCTTTTACGCAATTTCTCAGCCCTTGTAAACATTAAGAAATTCATCTGAGAGAA 1137
QY 249 ACCATTCCATTTGTCAGTGTGTAAAGGGTTAATTCTCGACAACTTGAAGAGACA 308
DB 1138 ACCCTTCAAAATGTAAAGATGTGGCAAGCTTTTACCAATTCCTCAACCTTCTAATCA 1197
QY 309 TGAATACCCATACAAAGTCAATTTAATGTATCTTGAATAATGTCAAGAAACATTTTA 368
DB 1198 TAGATTAATCTATCTGAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTA 1257
QY 369 TAAACATCAATC---TTTAAAGACATCATATATATCTGTTGATGAAAAAACAATTACGT 425
DB 1258 GCGACTCTCAACCTTACTTAACATTAATAATTAACATGTGCGAGAGAAACCTTCACAANT 1317
QY 426 TAAACATGTATTAAGTTTTCATGCACTTCAAAATTAGCAACATTAATAAACA 485
DB 1318 TGAAGATGTGGCAAAAGCTTTTAAATGATCTTCAAAATCTTACTATACATAGTTTATCA 1377
QY 486 TCATGTGTGATCTCCT 501
DB 1378 TACTGAGAGAAACCT 1393

RESULT 13
US-10-342-887-910
; Sequence 910, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernades, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
; LENGTH: 3839
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
US-10-342-887-910
Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAGATATATTTGACATATGAGGGGTGTGATTAAGCCCTATATGACCATCATATTATTA 128
DB AGAGAAACCTTACAAATGTGAAGAAATGTGCAAAAGCTTTAGCCATTTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCAGATATGATGACCGTATTAATGACAGTGAAGATTTGA 188
DB 1024 TAAACATTAAGAGATTCATCTGAGAGAAACCTTACAAATGT-----GAAGATGTG 1077
QY 189 TAAAGCATTTTTCAGAAAATTCACATTTGGAACAACATTTGTATCATTTCCGAAAAA 248
DB 1078 CAAAGCTTTTGAAGCATTTCTTCAAGCCCTTGTCTAAACATTAAGAAATTCATCTGAGAGAA 1137
QY 249 ACCATTCATTTGTCAGTGTGTGTAAAGGGTTAAATTCGACAAACATTTGAAAAAGA 308
DB 1138 ACCCTCAAAATGTAAAGAAATGTGCAAAAGCTTTTACAAATTCCTCAACCTTGTATCA 1197
QY 309 TGAATACACCCATACAAAGTCATTTAAATGTATCTGTTCAAGAAAAACATTAACGTG 425
DB 1198 TAAGATTAATCTCATCTGAAGAGAAACCTTACAAATGTAAAGATGTGACAAACCTTTTAA 1257
QY 369 TAAACATCATC---TTTAAAGACATCATATATATCTGTTCAAGAAAAACATTAACGTG 425
DB 1258 GGCATCTCAACCTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1317
QY 426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTAAGCAACATTAATTAAGAA 485
DB 1318 TGAAGATGTGGCAAGCTTTTAAATGATCTTCAATCTTACTATATAGTTATTTGA 1377
QY 486 TCATGTGATCTCCT 501
DB 1378 TACTGTGAGAAACCT 1393

RESULT 14
US-10-098-841-48
; Sequence 48, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
```

```

; SOFTWARE: pf_FL_genes Version 1.0
; SEQ ID NO 48
; LENGTH: 3502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(2583)
; NAME/KEY: misc_feature
; LOCATION: (1)..(3502)
; OTHER INFORMATION: n = a, t, c or g
US-10-098-841-48
Query Match      7.2%; Score 89; DB 13; Length 3502;
Best Local Similarity 51.7%; Pred. No. 4.8e-07;
Matches 256; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 79 TGCATATGAAGGGTGTGATTAAGCCATATATGACCATCTTTATGAGCAACATTTA 138
DB 2101 TACAAATGTGAAGATGTGGCAAAAGCTTTTAAAGTGTCTCAACCTTATCAATTAAG 2160
QY 139 AGAACCCAGATATGATGACCGATTAATGTACAGTGAAGATTTGATTAAGCATTT 198
DB 2161 ATTAATCTATCTGAGAGAAACCTTACAAATGT-----GAAGATGTGGCAAGCTTTT 2214
QY 199 TTCAGAAATTCACATTTGAAAACACATATTTGATCATTCGAAAAAAACCATTCAT 258
DB 2215 AAAGTGTCTCAACCTTTTACATACATTAATTAATTAATTAATTAATTAATTAATTA 2274
QY 259 TGTTCAGTGTGTGTAAAGGGGTTAATTTCTGACAAACCTTGAAGAAAGCATTAATCACC 318
DB 2275 TGTGAAAATGTGGCAAGCTTTTAAACGACCTCAACCTTATTTGAACATTAAGAAAT 2334
QY 319 CATACAAAGTCATTAATTAATGTACATTTGAAAATTTGCAAGAAACATTT---TATTAACAT 375
DB 2335 CATACTGAGAGCAACCTTACAAATGTGAAGATGTGGCAAGCATTTTATCAATTCCTCA 2394
QY 376 CATCTTTAAGACATCATATATATATCTGTTCAAGAAAAACATTAACGTGTAACATGT 435
DB 2395 CACCTTAATTAACATTAAGATTAATCTAATAAGAGCAACCTTACAAATGTAAAGATGT 2454
QY 436 AATAAAGTTTCACTGCACTTCAAAATTAAGCAACATTAATTAATTAATTAATTAATTA 495
DB 2455 GGCAAAGCTTCAACCAATTAATTAACCTTACTACACATTAACAAATTCATCTGAGAG 2514
QY 496 TCTCCTGTATCATATGATGATCATCCGTGTTTAAAAATTCCAAACCTTGTCAGTA 555
DB 2515 AAACCTTACAAACCTGAAAGATGTGACATGATTTTGAACACCTCAACCTTTTCAAAAC 2574
QY 556 TTACATTTTCATATA 570
DB 2575 ATTAATAATTAATA 2589

RESULT 15
US-10-029-386-25074
; Sequence 25074, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25074
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; OTHER INFORMATION: MAP TO AC011467.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
; OTHER INFORMATION: NT HIT: g115309152, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P35789, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AUI43734.1, EVALUE 0.00e+00
US-10-029-386-25074
```

```
Query Match 7.2%; Score 88.6; DB 16; Length 1466;
Best Local Similarity 53.3%; Pred. No. 4.1e-07;
Matches 237; Conservative 0; Mismatches 199; Indels 9; Gaps 2;
```

```
QY 60 ACSTCCCAAAAGTATATTTGCAATATGAGGGTGTGATTAAGCCTATATGACCATC 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 ACATTAAGTGAAGAAATCTTCAATATGTAAGATGTAAGAAAGTCATTTGCAATGCTTC 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACGATTAATGTAAGTGA 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 ACACCTTAGCTCAACATTAAGAAATGATGAGAGAAACCTCAAAATGTAA----- 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 CGATTGTGTAAGCATTTTTCAGAAATGACATTTGAAACACATATTTGATCAGATTC 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 -GATGTGGGAAAGCCTATATATGAGACCTCAACCTTTCTACATTAAGAAATTCATAC 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 240 CGAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACACACTT 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 TGGAAAGAAACCTTACAAATGCGAAGATGTGAAAAAGCCTTAAACGGCTCTCACACT 505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 300 GAAAAACATGAATACCCCATACAAAGTCATTAAATGTACATTTGAAAAATTGTCAAG 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 TACTACACATTAAGATATTCATCTGGAAGAAACCTCAAAATGTGAGAGGTGGCAA 565
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 360 AGCATTTTAATTAAT---CAATCTTAAGACATCATATTTATCTGTTGATGAAAAAC 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 AGCTTTTAACCAATCTGCAAAACCTTACTACATTAAGAAATTCATACCTGAGAAACC 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 417 ATTACGTGTAACATGTAATTAAGTTTCACTGCACTTCAAAATTAGCACACATTA 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 CTACAAATGTGAAGAAATGTGCGAGAGCTTTAGCCAGCTCTCAACCTTACTGCACTTA 685
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 477 ATTAAACATCATGTGATCTCT 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 GATAATTCATGCTGAGAGAAACCT 710
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: February 9, 2005, 11:50:30  
Job time : 2157.88 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:45:39 / Search time 3806.38 Seconds  
(without alignments)  
12360.150 Million cell updates/sec

Title: US-09-831-804-1\_COPY\_720\_1955  
Perfect score: 1236  
Sequence: 1 atgagtcgaagtcgaagcgaac.....aaacatcagtgattcttcga 1236

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	182.2	14.7	628	BZ298000	CG3957.f1
C 2	167.6	13.6	959	BZ298000	AL405671 T7 end of
C 3	103.8	8.4	801	CN759437	ID0AAA25C
4	103.8	8.4	898	CN761003	ID0AAA25C
5	98.2	7.9	1609	AK032220	Mus muscu
6	97.8	7.9	2647	BC047646	Homo sapi
7	97.6	7.9	745	AU123448	Homo sapi
8	93.4	7.6	727	CO434402	UT-W-HXO-
9	92.4	7.5	631	CN352904	CN352904 Homo sapi
10	92.4	7.5	2672	BC036394	CN352904 Homo sapi
11	92	7.4	591	AQ005136	CIT-HSP-2
12	92	7.4	687	CN362485	170004241
13	92	7.4	752	CN417582	170004240
14	91	7.4	593	B93867	CIT-HSP-228
15	91	7.4	820	BX414627	BX414627
16	90.8	7.3	2694	CN297781	170005321
17	90.4	7.3	2694	BC022527	Homo sapi
18	90.4	7.3	2698	BC037782	Homo sapi
19	90.2	7.3	2253	BC028252	Mus muscu
20	89.8	7.3	550	BP292533	BP292533
21	89.2	7.2	582	BP234906	BP234906
22	89.2	7.2	604	BE161630	MR3-HT044
23	89.2	7.2	638	CN341035	CN341035
24	89.2	7.2	673	CN353930	CN353930

25	89.2	7.2	3609	BC032590	BC032590 Homo sapi
26	89	7.2	2856	HSMB03730	AL832422 Homo sapi
27	88.6	7.2	535	AL705393	AL705393 DKFZ686M
28	88.6	7.2	612	BP238960	BP238960 BP238960
29	88.6	7.2	2174	AK033001	AK033001 Mus muscu
30	88.6	7.2	2330	BC037426	BC037426 Homo sapi
31	88.6	7.2	2622	AK033958	AK033958 Mus muscu
32	88.4	7.2	998	BX456765	BX456765
33	88	7.1	895	BU195416	BU195416 AGENCOURT
34	87.8	7.1	608	CN266758	CN266758 170004554
35	87.8	7.1	4251	CR749856	CR749856 Homo sapi
36	87.6	7.1	744	BX437291	BX437291 BX437291
37	87.2	7.1	579	AQ347265	AQ347265 RPI11-11
38	87.2	7.1	687	AQ389266	AQ389266 RPI11-15
39	87.2	7.1	907	BO423752	BO423752 AGENCOURT
40	87	7.0	660	CD770140	CD770140 AGENCOURT
41	87	7.0	842	BU507408	BU507408 AGENCOURT
42	86.8	7.0	521	AQ475626	AQ475626 CITBI-E1-
43	86.8	7.0	571	CN280110	CN280110 170006001
44	86.8	7.0	705	AQ194282	AQ194282 RPI11-61
45	86.8	7.0	860	AQ749175	AQ749175 HS_5575_A

#### ALIGNMENTS

RESULT 1  
BZ298000/c 628 bp DNA linear GSS 31-OCT-2002  
LOCUS CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata  
DEFINITION genomic clone CG3957, genomic survey sequence.  
ACCESSION BZ298000 GI:24440936  
VERSION BZ298000.1 GI:24440936  
KEYWORDS GSS.  
SOURCE Candida glabrata  
ORGANISM Candida glabrata  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
1 (bases 1 to 628)

REFERENCE  
Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.  
Evidence from comparative genomics for a complete sexual cycle in  
the 'asexual' pathogenic yeast Candida glabrata  
JOURNAL Genome Biol. 4 (2), R10 (2003)  
MEDLINE 22508158  
PubMed 12620120

#### COMMENT

Contact: Wong S  
Department of Genetics, Smurfit Institute  
Trinity College Dublin  
Dublin 2, Ireland  
Tel: 353 1 6082319  
Fax: 353 1 6798558  
Email: swong@tcd.ie  
Class: plasmid ends.  
Location/Qualifiers  
1..628  
/organism="Candida glabrata"  
/mol\_type="genomic DNA"  
/strain="CBS 138"  
/db\_xref="taxon:5478"  
/clone="CG3957"  
/clone\_1b="Candida glabrata Random Genomic Library"

#### FEATURES

source

#### ORIGIN

Query Match 14.7%, Score 182.2, DB 8, Length 628;  
Best Local Similarity 59.5%, Pred. No. 7.3e-26;  
Matches 348, Conservative 0, Mismatches 228, Indels 9, Gaps 2;

QY 84 ATTTGAAGGCTGATTAAGCCCTTATGACCATATTTAGACCAATTAAGAAC 143  
DB 625 ATTCGATTAATTTGGACAGGCGCTTTCAGAGCCGCTTCTACACACAGATAC 566  
QY 144 CCACAGTAATGATCGACCGCTAATATGTACAGTGACGATTTGATTAAGCATTTTTCAG 203





DEFINITION ID0AAA25CC01RMI ApMS Acyrthosiphon pisum cDNA clone ID0AAA25CC01 5', mRNA sequence.

ACCESSION CN759437

VERSION CN759437.1 GI:47533360

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

AUTHORS 1 (bases 1 to 801)

Stern,D., Tagu,D. and Wincker,P.

An expressed sequence tags database for the pea aphid Acyrthosiphon pisum

TITLE Unpublished (2004)

JOURNAL Contact: D. Tagu

COMMENT INRA Rennes

UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.48.51.65

Fax: +33.2.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGAAACAGCTATGACC

Plate: 25 row: C column: 1.

FEATURES

source

1..801

Location/Qualifiers

/organism="Acyrthosiphon pisum"

/mol\_type="mRNA"

/cultivar="developmentstage"

/db\_xref="taxon:7029"

/clone="ID0AAA25CC01"

/tissue\_type="whole insect"

/dev\_stage="nymphs and adults (parthenogenetic females)"

/lab\_host="XLI-Blue"

/clone\_lib="ApMS"

/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ; Harvesting date: 01/06/1999 ; Stress date: no stress ; Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 801;

Best Local Similarity 51.5%; Pred. No. 2.6e-10;

Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

82 ACATATGAAGGGTGTAAAGCCTATATGACCATCATTTATAGCAACATTTAGA 141

8 AAATGATTTCTGTGATAAGAGTTTACATATATACAGTTTAAAGCAATCAAG 67

142 ACCACAGTATATGATGACCGTATTAATGTACAGTGACGTTTGTATTAAGCATTTTC 201

68 ACACATACCGGTGAAAAGCCGTATGATGT-----GATTACTGTGATCAAGCGTTTCT 121

202 AGAAATACATTTGAAAACACATATTTGTACATTTCCGAAAAAAAACCATTCATTTGT 261

122 TTGAATACATTTAATAGCCATTAATAATGACATACCGCGCAAAAGCCCTTAATAATGT 181

262 TCAGTGTGTGTAAGGGGTTAATTTCCGACAAACATTTGAAAAGACATGAAATACCAT 321

182 GATTAAGTGTATCAAGCGTTTCTTAAGAAATCAAAATTTAACAAGCATATTAAGACAT 241

322 ACAAGTCATTTAATGATGATTTGAAAATGTCAAGAGCATTTTAAACATCAATCT 381

Db 242 ACCGCGAAAAGCCGTATTAATGTGAAAAGTGAACCAAGCGTTTCTCAGAAATCAGAT 301

Qy 382 TTAAAGCATCATATATATCTGTTCAAT---GAAAAACATTAACGTGAACATGTAAT 438

Db 302 TTAATACGTATATTAAGACACATACCGGCAAAAGCCGTATTAATGTGATACGTGAT 361

Qy 439 AAAGTTTCACTGACCTTCAAAATTGACACACATTAATTAACATCATGCTGATCT 498

Db 362 CAAGCGTTTCTCAAAAAGTGTATTAATTAAGCATTAATTAAGACATACCGGTGAAG 421

Qy 499 CTGCTTATCATGTGATCATCTCGTGTGTTTAAAAATTTCCAAACTGTGTAGATTA 558

Db 422 C---CGTTTAAATGTGCTAATCTGTGATCAAGCGTTTATTCGAAAACCATTTAAAAGC 478

Qy 559 CAATTCATATTAACACATCGATCCAAAACCTTAATGTCCCTAATGTGGTAAAGTTGT 618

Db 479 CACTCAAGACACATACAGGCCAAAAGCCGTATTAATGTGATACATGTAACCAAGCGTT 538

Qy 619 GTTGGGAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTCCATCATCAAA 678

Db 539 TCTCAAAATCAAAATTTAACAAGCATTAACAAGACACATGCCGTGAAAAGCTTTAA 598

Qy 679 ATATGACCTTGTGATTA 695

Db 599 TGGGATTAAGTGTATTA 615

RESULT 4

CN761003 898 bp mRNA linear EST 20-MAY-2004

LOCUS ID0AAA2AE05RMI ApMS Acyrthosiphon pisum cDNA clone ID0AAA2AE05 5', mRNA sequence.

DEFINITION CN761003

ACCESSION CN761003.1 GI:47534926

VERSION EST.

KEYWORDS Acyrthosiphon pisum (pea aphid)

SOURCE Acyrthosiphon pisum

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

AUTHORS 1 (bases 1 to 898)

Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B., Stern,D., Tagu,D. and Wincker,P.

An expressed sequence tags database for the pea aphid Acyrthosiphon pisum

TITLE Unpublished (2004)

JOURNAL Contact: D. Tagu

COMMENT INRA Rennes

UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.48.51.65

Fax: +33.2.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGAAACAGCTATGACC

Plate: 2 row: E column: 5.

FEATURES

source

1..898

Location/Qualifiers

/organism="Acyrthosiphon pisum"

/mol\_type="mRNA"

/cultivar="developmentstage"

/db\_xref="taxon:7029"

/clone="ID0AAA2AE05"

/tissue\_type="whole insect"

/dev\_stage="nymphs and adults (parthenogenetic females)"

/lab\_host="XLI-Blue"

/clone\_lib="ApMS"

/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ;



/tissue\_type="olfactory brain"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_etage="adult"  
330..1223  
/note="unnamed protein product; MSZF3 (FRAGMENT) homolog  
[Mus musculus] (SPR|088251, evidence: FAST, 87.1%id,  
100%length, match=255)  
putative"

/codon\_start=1  
/protein\_id="BAC27766.1"  
/db\_xref="GI:26328051"  
/translation="MLETYRNLAAGVYNEDSNIEHCESRRRHGRHRTGKEPYE  
GIQYGAFLVHSLQMRKIIHTGEKRYKNCQCKAYSRHSILQIHRTSGEPEYCN  
OGKATGHSKIKIHVHTGKPYKNCQCKAPAHNLQKERTHTGKPYKCDQC  
NKAFLVETIQVHKHTHTGKPYKNCQCKAPRHSILVHKIHTGKPYKNCQCK  
ALAHSTLQVHRTGKPYKNCQCKAPANOSYFQVHRIHTGKPYKCDQCKAF  
VSSSDLRHRTGKPYKNCQCKAPANOSYFQVHRIHTGKPYKCDQCKAF  
1594..1599  
/note="putative"  
/note="putative"  
1609  
/note="putative"

## ORIGIN

Query Match 7.9%; Score 98.2; DB 3; Length 1609;  
Best Local Similarity 51.2%; Pred. No. 3.6e-09;  
Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

60 ACCTCCCAAAAGTATATTTGACATATGAAAGGTGTATTAAGCCTTAATGACCAATC 119  
DB ACATCTGTGAGAGAAACGCTACAAATGCAATCAATGATGAAGCCTTAATGACGACAG 580  
QY 120 ATTATAGACCAATTTAGAACCCAGTATGATGCGCTTAATTAATGTCAGTGA 179  
DB 581 CATTCTACAAATCATTAAGAAACACATGATGAGAGAAACCTTAATGAATGA-----A 634  
QY 180 CGATTGTGATTAAGCAATTTTTCAGAAATATCAATTTGAAACACATATTTGATCAATTC 239  
DB 635 TCATGTGTGTAAGACCTTTACACACACATCTCATCTCAAAATACATATGTTACACATAC 694  
QY 240 CGAAAAAAAACCAATTCATTTGCTAGTGTGTGTAAGGGGTTAAATTCGACACACATTT 299  
DB 695 TGAGAGAGAAACCTTACAAATGTATCAATGTGTGAAGCCTTGCATGTATTAATTA 754  
QY 300 GAAAAACATGAATATCAACCATCAAGTCAATTTAATGATACATTTGAAAATGTCACAGA 359  
DB 755 TCAAAAAACATGAAGAAACACATCTGTGAGAGAAACCTTACAGTGTGATCAATGATATA 814  
QY 360 AGCAATTT---TATAACATCAATCTTTAAGACATCATATTTATCTGTTGATGAAAAAC 416  
DB 815 AGCCTTTGATATGAAGAGTATTTTACAGATTTCAAAAAACACATCTGTGAGAGAAAC 874  
QY 417 ATTACGTTTAAACAATGATTAAGTTCCTGCTGACCTTCAAAATTAGCACACATTA 476  
DB 875 TTACAAATGTAAGAAATGTGTAAGGCTTTGACACGACACAGTCATCTCAAAATGATTA 934  
QY 477 ATTAAAAACATCATGCTGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAA 536  
DB 935 AATAACACATACGCGAGAGAAAC---CTTACAAATGTAATCAATGTGTTAAAGCCCTTGC 991  
QY 537 TTTCCAAACCTTGGTCAGTATTACCAATTTCAATATTAACAACACTGCATCAAAACCTTAA 596  
DB 992 ATATCATATGACACTCCAGATACATCAAGAACACATCTGTGAGAGAGAAAGCCCTATGAT 1051  
QY 597 TCCTAAATGTGTAAAGGTGTGTGGAAGAAAGTTTATCTTCATCATATGTTAAATCA 656  
DB 1052 TGAGCAATGTGCGAAACCTTTGCAATCAAAAGTATTTCAAGTACATTAAGAAATACA 1111  
QY 657 TGATG 661  
DB 1112 TACTG 1116

RESULT 6

BC047646  
LOCUS BC047646 2647 bp mRNA linear HTC 19-NOV-2003  
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone  
IMAGE:4828290), with apparent retained intron.  
ACCESSION BC047646  
VERSION BC047646.1 GI:29126804  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

1 (bases 1 to 2647)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schler, G.D.,  
Altechul, S.F., Zeeberg, B., Buettow, K.H., Schefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carinci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schultz, J., Myers, R.M.,  
Butcherfield, Y.S., Krzywinski, M.I., Skalek, U., Small, D.E.,  
Scherer, A., Schein, J.E., Jones, J.S., and Maxam, A.M.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

## TITLE

JOURNAL MEDLINE  
PUBMED 12477932  
2 (bases 1 to 2647)  
Strausberg, R.  
Direct Submission  
Submitted (03-MAR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

## JOURNAL

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [medpaxil.stanford.edu](mailto:medpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

## FEATURES

## source

1..2647  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4828290"  
/tissue\_type="Testis"  
/clone\_lib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

## ORIGIN

Query Match

7.9%; Score 97.8; DB 3; Length 2647;

Best Local Similarity 50.1%; Pred. No. 4.4e-09;  
Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;

```
QY 7 GAAAGTACGAAACCAATGATATCATCTTTATATATCTCTCTCTTCATCAGCTCC 66
    |||||
Db 1235 GAATGGGAAACCTTTAACGATCTCTCACTCTACTACCATTAAGATTAATCTACT 1294
QY 67 AAAAGTATATTTGCAATATGAAAGGTGATTAAGGCTTAATATGACCATTAATTA 126
    |||||
Db 1295 GGAGAGAAACCCCTACAAATGTAAGATGTGTAAGCTTTTAACGATCTCAACCTT 1354
QY 127 GAGCAACATTTAGAACCCACATTAATGATCGACCTTAATTAATGATGAGACGATTT 186
    |||||
Db 1355 ACTACCATGAAATAATCTACTGAGAGAAACCTTACAAAGT-----GAAAGATGT 1408
QY 187 GATTAAGCATTTTTCAGAAATCATATTTGAAACATATTTGATCATCTCCGAAAA 246
    |||||
Db 1409 GCGAAAGCCTTTAGAGAGCTCTCAACCTTACTACATAGATTAATCTACTGAGAG 1468
QY 247 AAACATTCATTTTCACTGATGTGTGTAAGGGGTTAATTTCTGACAACTTGAAAA 306
    |||||
Db 1469 AAACCTTACAAATGTAATAAATGTGAAAAAGCCTTTAACAGCTGACACCTTACACA 1528
QY 307 CATGAATACCCATACAAAGTCATTTAATGATCATTTGAAATTTGTCAAGAGCATTT 366
    |||||
Db 1529 CATGAGTATTTATCTGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCTTT 1588
QY 367 TATTA--ACATCAATCTTTAAGACATCATATATATATCTGATGAAAAACATTAAG 423
    |||||
Db 1589 AATCATTTCTCACACCTTACTACATAGATTAATCTACTGAGAGAAACCTTACAAA 1648
QY 424 TGTAAACATGTAATAAGTTTCACTGACCTTCAAAATGACACACATTAATTAATA 483
    |||||
Db 1649 TGTAAAGATGTGTAAAGCTTTTAACACTCTTCAACCTTACTAATCATTAAGATTA 1708
QY 484 CATCATGTGTGATCTCTGCTTATCATATGATCATCTGCTGTTTAAAAATTTCCAA 543
    |||||
Db 1709 CATACGTGAGAGAGAGCTTACAAATGTAAAGATGTGAAAAAGCTTTTAACAACTCA 1768
QY 544 ACTGTGTCATTTTACATTTTCAATTAATAACAATGATCAACCAATTAATGCTTAA 603
    |||||
Db 1769 AAACCTTACTGACATTAAGAAAAATTCATCTGAGAGAAACCTTA--TGAATGTAAAA 1825
QY 604 TGTGTAAAGTTGTGTGGAAAAAAGTTTATCTTCAATATGTTAATGTCATGATGA 662
    |||||
Db 1826 TGTGGCAAGCTTTTAACGAGCTCTCAATCTTACTAGACATTAAGAAAAAGTCAACGA 1884
```

RESULT 7  
AUI23448 745 bp mRNA linear EST 01-AUG-2002  
LOCUS AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA  
DEFINITION sequence.

ACCESSION AUI23448  
VERSION AUI23448.1 GI:10948164  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 745)  
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,  
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and  
Isegai, T.

TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,  
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,  
Suzuki, Y., Sugano, S., Isegai, T.)  
Unpublished (2000)  
CONTACT: Takao Isegai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975

Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

## FEATURES

source  
1..745  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RM2000319"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/clone\_id="NT2RM2"  
/note="Vector: pME18PFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

## ORIGIN

Query Match 7.9%; Score 97.6; DB 1; Length 745;  
Best Local Similarity 50.2%; Pred. No. 4.4e-09;  
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

```
QY 7 GAAAGTACGAAACCAATGATATCATCTTTATATATCTTCTTCTTCATCAGCTCC 66
    |||||
Db 87 GAATGGGAAACCTTTAACGATCTCTCACTCTTACTACCATTAAGATTAATCTACT 146
QY 67 AAAAGTATATTTGCAATATGAAAGGTGATTAAGGCTTAATATGACCATTAATTA 126
    |||||
Db 147 GAGAGAAACCTTACAAATGTAAGATGTGTAAGCTTTTAACGCTTCTTCAACCTT 206
QY 127 GAGCAACATTTAAGAACCCACATATGATCGACCTTAATATGATGAGACGATTT 186
    |||||
Db 207 ACTACCATGAAATAATCTACTGAGAGAAACCTTCAAAATGT-----GAAAGATGT 260
QY 187 GATTAAGCATTTTTCAGAAATCATATTTGAAACATATTTGTACATTTCCGAAAA 246
    |||||
Db 261 GCGAAAGCCTTTAGAGAGCTCTCAACCTTACTACATAGATTAATCTACTGAGAG 320
QY 247 AAACATTCATTTTCACTGATGTGTGTAAGGGGTTAATTTCTGACAACTTGAAAA 306
    |||||
Db 321 AAACCTTACAAATGTAATAAATGTGAAAAAGCTTTAACAGCTGACACCTTACTACA 380
QY 307 CATGAATACCCATACAAAGTCATTTAATGATCAATTTGAAATTTGTCAAGAGCATTT 366
    |||||
Db 381 CATGAGTATTTATCTACTGAGAGAAACCTTCAAAATGTGAAAAATGTGAAAAAGCTTT 440
QY 367 TATTA--ACATCAATCTTTAAGACATCATATATATCTGTTCAAGAAAAACATTAAG 423
    |||||
Db 441 AATCATTTTCTCACACCTTACTACATAGATTAATCTACTGAGAGAAACCTTACAAA 500
QY 424 TGTAAACATGTAATAAGTTTTCAGACCTTCAAAATGACACATTAATTAATAA 483
    |||||
Db 501 TGTAAAGATGTGTAAAGCTTTTAACACTCTTCAACCTTACTAATCATTAAGATTA 560
QY 484 CATCATGTGTGATCTCTGCTTATCATATGATGATCATCTGCTGTTTAAAAATTTCCA 543
    |||||
Db 561 CATACTGAGAGAAACCTTACAAATGTAAGATGTGAAAAAGCTTTTAACCAATCTCA 620
QY 544 ACTGTGTCATTTTCAATTTTCAATTAATAACAATGATCAACCAATTAATGCTTAA 603
    |||||
Db 621 AAACCTTACTGACATTAAGAAAAATTCATCTGAGAGAAACCTTA--TGAATGTAAAA 677
QY 604 TGTGTAAAGTTGTGTGGAAAAAAGTTTATCTTCAATATGTTAATGTCAT 657
    |||||
Db 678 TGTGGCAAGCTTTTAACGAGCTCTCAATCTTACTAGACATTAAGAAAAAGTCA 731
```

RESULT 8  
CO434402 727 bp mRNA linear EST 06-JUL-2004  
LOCUS CO434402  
DEFINITION UT-M-HX0-csl-h-09-0-UT.r1 NIH\_BMAP\_HX0 Mus musculus cDNA clone  
IMAGE:30687440 5', mRNA sequence.

ACCESSION CO434402  
 VERSION CO434402.1 GI:49680696  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 727)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Dr. James Ian University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at <http://genome.iowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
 Seq primer: pyx-5.  
 Location/Qualifiers  
 1. 727  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30687440"  
 /rname\_type="whole eye"  
 /dev\_stage="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_id="N1H\_BMAP\_HX0"  
 /note="Organ: Eye; Vector: pyx-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATGATTACG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN  
 Query Match 7.6%; Score 93.4; DB 7; Length 727;  
 Best Local Similarity 52.1%; Pred. No. 3e-08;  
 Matches 261; Conservative 0; Mismatches 231; Indels 9; Gaps 2;  
 4 AGTGAAGTGAACGAAACCAATGATATCATCTTATATCTCTTCTGATACGCT 63  
 DB AATCAGTGTATTAAGCTTTCTCGATACGATCTTCAATCTCAATGAAGACAT 137  
 64 CCCAAAAAGTATTTGACATATGAAGGTGTGATTAAGCCTATATCGACCATCTTA 123  
 DB ACTGGAGAGAAACCTTCAATGTATCAGTGTGATTAAGCCTTTTGTGAAGATGAGT 197  
 124 TTGAGGACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTATACAGTGACGAT 183  
 DB CTCGAAAACCATGTAAAGATACATACGTGAGAGAAACCTTCAAAATGTA-----ATCAA 251  
 184 TGTGATAAGCATTTTCAGAAATTCACATTTGGAACACATATTTGATTCATTCGGA 243  
 DB TGTGATTAAGCTTTTCCCATATACCTCCACATTTCAATGAGAGAACATACCGGA 311  
 244 AAAAACCATTTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGCACAACTTGAA 303  
 DB 312 GAGAAACCTTAATGTATATGATGTATTAAGCCTTTCCCATATATCTACACCTCCAC 371

ORIGIN  
 Query Match 7.5%; Score 92.4; DB 7; Length 631;  
 Best Local Similarity 54.1%; Pred. No. 4.7e-08;  
 Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;  
 69 AAGTATATTTGACATATGAAGGTGTGATTAAGCCTATATCGACCATATTAATGA 128  
 DB AGAGAAAGCCTTAAAGTATGAAGAAAGTGAAGAAAGCTTTTAAACCAATCTCAAAACTTAC 142  
 129 GCAACATTTTAAGAACCCACAGTATGATCGACCGTATTAATGTATACAGTGACGATGTGA 188  
 DB TGAACATTAAGAAATTCATATCTGAGAGAAACCTTATGATGT-----GAAAAATGTGG 196  
 189 TAAAGCATTTTTCAGAAATTCACATTTGGAACACATATTTGATTCACATTTCCGAAAAAA 248  
 DB 197 CAAAGCTTTTAACCAAGTCTCAATCTTACTAGACATTAAGAAAGTCATACAGAGAGAA 256

FEATURES  
 source  
 1. 631  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /rname\_type="embryonic stem cells, DMSO-treated H9 cell line"  
 /note="oligo dT primed, full-length enriched cDNA library from DMSO-treated hES cell line H9 (p22) maintained in feeder-free conditions"  
 Location/Qualifiers  
 RESULT 9  
 CN352904 631 bp mRNA linear EST 16-MAY-2004  
 LOCUS CN352904 17000600057466 GRN\_PREHEP Homo sapiens cDNA 5', mRNA sequence.  
 DEFINITION CN352904  
 ACCESSION CN352904  
 VERSION CN352904.1 GI:47352838  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 631)  
 Brandenberger R., Wei, H., Zhang, S., Lei, S., Muraie, J., Flak, G. J., Li, Y., Xu, C., Fan, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.  
 Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation  
 Nat. Biotechnol. 22 (6), 707-716 (2004)  
 Contact: Brandenberger R  
 Regenerative Medicine  
 Geron Corporation  
 230 Constitution Drive, Menlo Park, CA 94025, USA  
 Tel: 650 473 8658  
 Fax: 650 473 7760  
 Email: [rbrandenberger@geron.com](mailto:rbrandenberger@geron.com)  
 Insert Length: 631 Std Error: 0.00.  
 Location/Qualifiers

QY 249 ACCATTCATTGTCAGTGTGTGTAAGGGGTTAATCTCGAACACCTTGAAAGACA 308  
 Db 257 ACCTTACAAATGTGAAGATGTGGCAAGGTTTAAATGCGCCCTCAACCTTACTATCA 316  
 QY 309 TGAATACCCATACAAAGTCAATTTAAATGTACATTGTGAATGTGCAAGAACTTTTA 368  
 Db 317 TAAAGATTAATTCATCTGAGAGAGAAACCATACAAATGTGAAGATGTGGCAAGCTTTTA 376  
 QY 369 TAAA---CATCATCTTTTAAGACATCATATATATCTGTTTCATGAAAAAATTAACGTG 425  
 Db 377 CCAATCTCAAAACTTACCAACATTAAGAAATTCATCTACTGAGAGAAACCTTACACATG 436  
 QY 426 TAAACATGTATTAATTAAGTTTCACTGACCTTCAAAAATTAGCAACAATTAATTAACA 485  
 Db 437 TGAAGATGTGGCAAGGCTTTAACAGTCTCAAACTTACTTAACATTAAGAAATTCGA 496  
 QY 486 TCATGTGTGATCTCCT 501  
 Db 497 TACTGGAGAAAAAACCCT 512

RESULT 10  
 BC036394 2672 bp mRNA linear HTC 19-NOV-2003  
 LOCUS  
 DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTP1), mRNA (cDNA clone  
 IMAGE:5259399), containing frame-shift errors.  
 ACCESSION BC036394  
 VERSION BC036394.1 GI:23025784  
 KEYWORDS HTC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 2672)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stalcron, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Wooley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Grevenhock, Y.,  
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schultz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,  
 Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE  
 PUBLISHED 22388257  
 REFERENCE 12477932  
 2 (bases 1 to 2672)  
 Strausberg, R.  
 Direct Submision  
 Submitted (05-AUG-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs-rt@mail.nih.gov](mailto:cgabbs-rt@mail.nih.gov)  
 Tissue Procurement: Niklos Palkovits, M.D., Ph.D.,  
 CDNA Library Preparation: Michael J. Bronshtein (NNGRI) & Shiroki  
 Toshitaki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.W.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systembiology.org>

contact: [amadane@systembiology.org](mailto:amadane@systembiology.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
 Clone distribution: MGC clone distribution information can be found  
 through the I.W.A.G.E. Consortium/LINL at: <http://image.linl.gov>  
 Series: IRAP Plate: 73 Row: a Column: 16  
 This clone has the following problem: frame shifted.  
 Location/Qualifiers  
 1..2672  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5259399"  
 /issue\_type="Brain, hippocampus"  
 /clone\_id="NTH MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

ORIGIN  
 Query Match 7.5% Score 92.4; DB 3; Length 2672;  
 Best Local Similarity 54.1%; Pred. No. 5,2e-08;  
 Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;  
 QY 69 AAGATATATTTGCACATATGAAGGGTGTGATTAAGCCATATATCGACCATCATTTATGA 128  
 Db 1745 AGAGAAAGCTTACAAATGTAAAGATGTGAAGAAAGCTTTTAACCAATCCTCAAACTTAC 1804  
 QY 129 GCAATTTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAAGCATTTGCA 188  
 Db 1805 TGAAATTAAGAAATTTCTACTGAGAGAAACCTTATGAATGT-----GAAAAATGTGG 1858  
 QY 189 TAAAGCATTTTTCAGAAATCATTTGGAACATTTGATATTCATTCGATTCGAAAAAA 248  
 Db 1859 CAAGCTTTTAAACAGCTCTCAATCTTACTAGCATTAAGAAATCATTAAGAGAGAA 1918  
 QY 249 ACCATTCATTGTCAGTGTGTGTAAGGGGTTAATCTCGACCAACATTTGAAAAACA 308  
 Db 1919 ACCTTACAAATGTGAAGATGTGGCAAGGTTTAAATGGCCCTCAACCTTACTATCA 1978  
 QY 309 TGAATACCCATACAAAGTCAATTTAAATGTACATTGTGAATTTGCAAGAAAGCTTTTA 368  
 Db 1979 TAAAGTAAATTCATCTGAGAGAGAAACCATACAAATGTGAAGATGTGCAAGCTTTTAA 2038  
 QY 369 TAAA---CATCAATCTTTAAGACATCATATATATCTGTCATGAAAAAATTAACGTG 425  
 Db 2039 CCAATCTCAAAACTTACCAACATTAAGAAATTCATCTACTGAGAGAAACCTTACACATG 2098  
 QY 426 TAAACATGTATTAATTAAGTTTCACTCGACCTTCAAAAATTAGCAACAATTAATTAACA 485  
 Db 2099 TGAAGATGTGGCAAGGCTTTAACAGTCTCAAACTTACTTAACATTAAGAAATTCGA 2158  
 QY 486 TCATGTGTGATCTCCT 501  
 Db 2159 TACTGGAGAAAAAACCCT 2174

RESULT 11  
 A0005136 591 bp DNA linear GSS 27-JUN-1998  
 LOCUS  
 DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017,  
 genomic survey sequence.  
 ACCESSION A0005136  
 VERSION A0005136.1 GI:3082581  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 591)  
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,  
 Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,  
 Simon, M. and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)  
 JOURNAL Unpublished (1998)  
 COMMENT Other GSSs: CIT-HSP-2290017.TR  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdamas@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tcdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: M13-21,  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..591  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7150844"  
 /db\_xref="taxon:9606"  
 /clone="2290017"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /clone\_lib="CIT-HSP"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

ORIGIN  
 Query Match 7.4%; Score 92; DB 8; Length 591;  
 Best Local Similarity 52.2%; Pred. No. 5.6e-08;  
 Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

D 7 GAAAGTGCAGAAACCAATGATATCATCTTAAATCTCTCTTCATCAGCTCC 66  
 69 GAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACACATAGATTAATTCATCT 128  
 QY AAAAAATATTTTGACATATGAAAGGTGTGATTAAGCCATATGACCATATATTA 126  
 129 GGAGAGAGGCCCTTCAATGTGAAGATGTGTAAGCTTTTAACCACTTCAGCCCTT 188  
 Db 127 GAGCAACATTTAAGAACCCACAGATATGATCGACCGTATTAATGTACAGCGATTGT 186  
 189 ACTACACATATGATTCATTCATGTTTAAGAAAAACCTCAAAATGT-----GAAATGT 242  
 QY 187 GATTAAGCATTTTTCAGAAAAATCACTTTGAAAAACATATTGTATCACTTCGAAAA 246  
 243 GACAAAGCTTTTAACGATTCCTATACCTTAACATAAGATAATTCCTGGAGAG 302  
 Db 247 AAACCATTCATTTGTAGTGTGTGTGTAAGGGGTTAATTCGACAACCTTGAAAA 306  
 303 AAATCTTAACAATGTGAACATGTGGCAAGGCTTTAACTGTCTTCAACCTTACAAA 362  
 QY 307 CATGAATTCACCAATCAAAAGTCATTTAAATGTACATTTGAAAAATGTCAAGAGATT 366  
 363 CATTAAGAAATTCCTACTGAGAGAAACCTCAAAATGTAAAGATGTGGCAAGCTTT 422  
 Db 367 TATTAACAT--CAATCTTAAAGACATCATATATATCTGTGATGAAAAACATTAA 423  
 423 AATGTGCTTCAACCTTACTACACATTAAGATGATTCATCTGGAAGAAACCTTA 482  
 QY 424 TGTAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTAAGACAACATTAATTA 483  
 483 TGTGAAGAATGTGGCAAGGCTTTAACACCTCTCAAACTTACTATACATAAGATAAT 542  
 Db 484 CATCATGTG 493  
 543 CATCTGGAG 552

LOCUS CN362485 687 bp mRNA linear EST 16-MAY-2004  
 DEFINITION 17000424188118 GRN\_EB Homo sapiens cDNA 5', mRNA sequence.  
 ACCESSION CN362485  
 VERSION CN362485.1 GI:47362419  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 687)  
 Brandenberger,R., Wei,H., Zhang,S., Lei,S., Muzage,J., Fisk,G.J., Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R., Lebkowski,J and Stanton,L.W.  
 Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation  
 Nat. Biotechnol. 22 (6), 707-716 (2004)  
 CONTACT: Brandenberger R  
 Regenerative Medicine  
 Geron Corporation  
 230 Constitution Drive, Menlo Park, CA 94025, USA  
 Tel: 650 473 8658  
 Fax: 650 473 7760  
 Email: rbrandenberger@geron.com  
 Insert length: 687 Std Error: 0.00.  
 Location/Qualifiers  
 1..687  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /issue\_type="embryonic stem cells, embryoid bodies derived from H1, H7 and H9 cells"  
 /clone\_lib="GRN EB"  
 /note="Toiigo dT primed, full-length enriched cDNA library from embryoid body outgrowth derived from hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions."

ORIGIN  
 Query Match 7.4%; Score 92; DB 7; Length 687;  
 Best Local Similarity 52.2%; Pred. No. 5.6e-08;  
 Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

D 7 GAAAGTGCAGAAACCAATGATATCATCTTAAATCTCTCTTCATCAGCTCC 66  
 69 GAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACACATAGATTAATTCATCT 128  
 QY AAAAAATATTTTGACATATGAAAGGTGTGATTAAGCCATATGACCATATATTA 126  
 183 GAATGTGGCAAGCCTTTCATGATGCTTTCACACCTTACACATAGATTAATTCATCT 242  
 Db 243 AGATGAATTTTTCACAAATGTGAAGCATATGGAAGGCTTTAACTGTCTCAACCTT 302  
 QY 127 GAGCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGCGATTGT 186  
 303 AAATTAACAATGAAGATTCATCTGAGAAAAACCTTAACAATGTAAA-----GAATGT 356  
 QY 187 GATTAAGCATTTTTCAGAAAAATCACTTTGAAAAACATATTGTATCACTTCGAAAA 246  
 357 GGCAAGGCTTTTACACAGCTTCACCTTATTTAGCATTAAGAAATTCCTACTGAAG 416  
 Db 247 AAACCATTCATTTGTAGTGTGTGTGTAAGGGGTTAATTCGACAACCTTGAAAA 306  
 417 AAACCTTAACAATGTGAAGATGTGGCAAGGCTTTTAACCAAGTATGACCTTACTA 476  
 QY 307 CATGAATTCACCAATCAAAAGTCATTTAAATGTACATTTGAAAAATGTCAAGAGATT 366  
 477 CATTAATTAATTCATCTGAGGAAATTCCTCAAAATGTGAAGAAATGTGAGAGCTTT 536  
 Db 367 TATTAACATCAATCTTAAAGACATCATATATATCTGT--CATGAATAACATTAA 423  
 537 AACCAAGCTCAAGGCTTACTGAACATTAAGTTATTCATACCGGAGAGAACTTATGA 596  
 QY 424 TGTAAACAATGTAAAGTTTTCATCTGACCTTCAAAATTAAGACAACATTAATTA 483  
 597 TGTGAAGAATGTGGCAAGGCTTTTAACGATCTCTCAAACTTACTGAACATTAAGTACAT 656



QY 484 CATCATGTG 493  
DB 657 CATACTGAG 666

RESULT 13  
CN417582  
LOCUS CN417582 752 bp mRNA linear EST 16-MAY-2004  
DEFINITION 17000424010604 GRN\_EB Homo sapiens cDNA 5', mRNA sequence.  
ACCESSION CN417582  
VERSION CN417582.1 GI:47405176  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 752)  
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,  
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,  
Lebkowski, J. and Stanton, L.W.  
Transcriptome characterization elucidates signaling networks that  
control human ES cell growth and differentiation  
Nat. Biotechnol. 22 (6), 707-716 (2004)  
Contact: Brandenberger R  
JOURNAL  
COMMENT  
Regenerative Medicine  
230 Conelution Drive, Menlo Park, CA 94025, USA  
Tel: 650 473 8658  
Fax: 650 473 7760  
Email: rbrandenberger@geron.com  
Insert Length: 752 Std Error: 0.00.  
Location/Qualifiers  
1..752  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="embryonic stem cells, embryoid bodies  
derived from H1, H7 and H9 cells"  
/clone\_id="GRN\_EB"  
/note="Toigo dt primed, full-length enriched cDNA library  
from embryoid body outgrowths derived from h9s cell lines  
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free  
conditions."

ORIGIN  
Query Match 7.4%; Score 92; DB 7; Length 752;  
Best Local Similarity 52.0%; Pred. No. 5.7e-08;  
Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATGATATCATCTTTAATATCTTCTTCATCAGGTC 66  
DB 72 GAATGGGAAAGCCTTACCATCTTACACCTTACATACATAGTAATCTACT 131  
QY 67 AAAAGTATATTTGACATATGAAGGGTGATTAAGCCTATPANTGACCATATTATTA 126  
DB 132 GGAGAGAAAGCCCTTCAATGTGAAGATGTGTAAGCTTTAACCACCTTCAGCCCTT 191  
QY 127 GAGCAACATTTAAGAACCCAGATATGATGACCTATTAATGTGACATGACATTT 186  
DB 192 GCTACACATTAAGCTTCTTATGTTAAAGAAAAACCTTACAAATGT-----GAAAGATGT 245  
QY 187 GATTAAGCATTTTTCAGAAATCACATTTGAAACACATATTTGATCATTCCTCGAAAAA 246  
DB 246 GACAAAGCTTTTAAACGATTTCTATCTTACTTAACATTAATTAATTTCTGAGAG 305  
QY 247 AAACATTTCCATTTGCTGCTGTGTGTAAGGGGTTAATTCGACAAACATTTGAAAGA 306  
DB 306 AAATCTTAACAATGTGAACAATGTGGCAAGGCTTTAATGCTTCAACCTTACAAAA 365  
QY 307 CATGAATATACCATTAAGATCATTTAATGTACATTTGAAATTTGTCAAGAAGATTT 366  
DB 366 CATAGAGAAATTCATATCTGAGAGAAACCTTCAATGTGAAGAAATGTGGCAAGCTTT 425

QY 367 TATAAACAAT---CAATCTTAAGACATATATATTCGTTCATGAAAAAATTAACG 423  
DB 426 AATGTCTCTTACACCTTACTACATAGATGATTAATCTGAGAGAAACCTTACAAA 485  
QY 424 TGTAAACAATGTAAATTAAGTTTCTGACCTTAAATTAAGCAACATTAATTAATAA 483  
DB 486 TGTGAAGATGTGGCAAGCCTTTAACCATCTTCAAACTTAATCATTAAGTAATTT 545  
QY 484 CATCATGTGTGATCTCCT 501  
DB 546 CATACTGAGAGAAACCT 563

RESULT 14  
B99387  
LOCUS B99387 593 bp DNA linear GSS 26-JUN-1998  
DEFINITION CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,  
genomic survey sequence.  
ACCESSION B99387  
VERSION B99387.1 GI:3027197  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 593)  
Adams, M.D., Rounnley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,  
Simon, M., and Venter, J.C., Granger, D., Suh, E., Wible, C., Shizuya, H.,  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
Unpublished (1998)  
Contact: Mark Adams  
JOURNAL  
COMMENT  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamas@tigr.org  
Clones are available from Research Genetics (info@reagen.com). BAC  
end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
source  
location/Qualifiers  
1..593  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7147044"  
/db\_xref="taxon:9606"  
/clone="2281A10"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_id="CIT-HSP"  
/note="Vector: pBelobac11; site\_1: HindIII; site\_2:  
HindIII"

ORIGIN  
Query Match 7.4%; Score 91; DB 8; Length 593;  
Best Local Similarity 54.4%; Pred. No. 8.8e-08;  
Matches 231; Conservative 0; Mismatches 185; Indels 9; Gaps 2;

QY 69 AAAGTATTTTGCATATGAAGGGTGATTAAGCCTATTAATGACCATGATTTTGA 128  
DB 47 AAAGAAACCTTAACAATGTGAAGATGTGGCAAGCTTTTACCGATCTTCAACCTTAG 106  
QY 129 GCAACATTTAAGAACCCAGATATGATGACCCGATTAATGTACAGTGAAGATTTGA 188  
DB 107 AAACATGAGTAATTAATCTGAGAGAAACCTTACAAATGT-----GAAAGATGTG 160  
QY 189 TAAACATTTTTCAGAAAAATCACATTTGAAACATATTTGTATCATTCGAAAAAAA 248



**This Page Blank (uspto)**

100.0%; Score 1239; DB 6; Length 1239;

Best Local Similarity 100.0%; Pred. No. 5.7e-190;  
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTATCA 60
Db 1 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTATCA 60
QY 61 CGTCCCAAAAAGTATTTTGACATATGAAGGGTGTGATTAAGCCATTAATCGACATCA 120
Db 61 CGTCCCAAAAAGTATTTTGACATATGAAGGGTGTGATTAAGCCATTAATCGACATCA 120
QY 121 TTATTAGCAACATTTAAGAACCCACAGTAATGACCCGATTAATAGTACAGTGCAC 180
Db 121 TTATTAGCAACATTTAAGAACCCACAGTAATGACCCGATTAATAGTACAGTGCAC 180
QY 181 GATTGTGATTAAGCATTTTTCAGAAATGACATTTGGAACAATATGATTCACATTC 240
Db 181 GATTGTGATTAAGCATTTTTCAGAAATGACATTTGGAACAATATGATTCACATTC 240
QY 241 GAAAAAACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACACACTTG 300
Db 241 GAAAAAACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACACACTTG 300
QY 301 AAAAGCATGAATTCACCAATCAAGTCAATTAATGTACATTTGAAAATTTGTCAAGA 360
Db 301 AAAAGCATGAATTCACCAATCAAGTCAATTAATGTACATTTGAAAATTTGTCAAGA 360
QY 361 GCATTTTAAACATCAATCTTTAAGACATCATATATTTCTGTCATGAAAAACATTA 420
Db 361 GCATTTTAAACATCAATCTTTAAGACATCATATATTTCTGTCATGAAAAACATTA 420
QY 421 ACGTGAACCATGTATTAAGTTTCTGTCGACCTTCAAAATTTAGCACACATAATTA 480
Db 421 ACGTGAACCATGTATTAAGTTTCTGTCGACCTTCAAAATTTAGCACACATAATTA 480
QY 481 AAAACATCATGTGATCTCTGCTTATCATATGATCATCTGCTGTTTAAAAATTTTC 540
Db 481 AAAACATCATGTGATCTCTGCTTATCATATGATCATCTGCTGTTTAAAAATTTTC 540
QY 541 CAAACTGTGTCATTTAATTTCAATTTCAATTAACCACTGATCCAAACCTTAAATGCT 600
Db 541 CAAACTGTGTCATTTAATTTCAATTTCAATTAACCACTGATCCAAACCTTAAATGCT 600
QY 601 AAATGTGTAAAGTTGTGTGGAAAAAGGTTATCTTCAATATGTTAAGTCATGAT 660
Db 601 AAATGTGTAAAGTTGTGTGGAAAAAGGTTATCTTCAATATGTTAAGTCATGAT 660
QY 661 GATTCTACATGATCAAAATATGACCTGTGATTAATGATGATGAGGAAATTTGCAAG 720
Db 661 GATTCTACATGATCAAAATATGACCTGTGATTAATGATGATGATGAGGAAATTTGCAAG 720
QY 721 AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGATTAATCCCTGATGATTA 780
Db 721 AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGATTAATCCCTGATGATTA 780
QY 781 TTAAGGAACCTGAAGTGAAGAAATTTAGGAACCTATTAAGATCAAGATCGAAATTAAT 840
Db 781 TTAAGGAACCTGAAGTGAAGAAATTTAGGAACCTATTAAGATCAAGATCGAAATTAAT 840
QY 841 AATTTCATGATTAAGAAACAGAAATTTAAAGTGAAGAAAGATGAAGATGAAGAA 900
Db 841 AATTTCATGATTAAGAAACAGAAATTTAAAGTGAAGAAAGATGAAGATGAAGAA 900
QY 901 GATAGCTAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 901 GATAGCTAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 ATTAATATCTTTACTGCTTCTTGAAGGTTCAAGAGTCTTAACTTAATTTCTGAAT 1020
Db 961 ATTAATATCTTTACTGCTTCTTGAAGGTTCAAGAGTCTTAACTTAATTTCTGAAT 1020
QY 1021 AGTGGAGAGATCAATTTGCTTAAGAAATTAATTTGATGAATGTTTCTAGAGAAAT 1080

```

Db 1021 AGTGGAGAGATCAATTTGCTTAAGAAATTAATTTGATGAATGTTTCTAGAGAAAT 1080

QY 1081 GATTACGTCGACATTTGAAATGCGATGATGATTAATTTAACAAGAAATGATCATTTCTTA 1140

Db 1081 GATTACGTCGACATTTGAAATGCGATGATGATTAATTTAACAAGAAATGATCATTTCTTA 1140

QY 1141 AATAGTATGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1200

Db 1141 AATAGTATGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1200

QY 1201 GATTATTGCCAATGAAGAAACATCACTGATTTCTCGATTA 1239

Db 1201 GATTATTGCCAATGAAGAAACATCACTGATTTCTCGATTA 1239

RESULT 2  
BD274349  
LOCUS BD274349 2060 bp DNA linear PAT 17-JUL-2003  
DEFINITION Candida albicans tfl1111 gene (Catf1111) and the coded CATP1111 protein.  
ACCESSION BD274349  
VERSION BD274349.1 GI:33084117  
KEYWORDS JP 2002531068-A/1.  
SOURCE Candida albicans  
ORGANISM Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
REFERENCE 1 (bases 1 to 2060)  
AUTHORS Pallier, F. B., Camier, S. and Sentenac, A.  
TITLE Candida albicans tfl1111 gene (Catf1111) and the coded CATP1111  
JOURNAL Patent: JP 2002531068-A 1 24-SEP-2002;  
COMMENT AVENTIS PHARMA SA  
OS Candida albicans  
PN JP 2002531068-A/1  
PD 24-SEP-2002  
PF 08-NOV-1999 JP 2000581204  
PR 10-NOV-1998 FR 98/14147  
PI FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC PC  
C12N15/09, C12N15/09, A61K45/00, A61P31/10, C07K14/40, PC  
C07K16/14,  
PC C12N1/19, C12N1/21, C12P21/02, C12P1/02, G01N33/15, G01N33/50, PC  
G01N33/53,  
PC G01N33/566, G01N33/569// (C12N15/09, C12P1/725), C12N15/00, C12N15/00,  
PC (C12N15/00, C12P1/725)  
CC Candida albicans tfl1111 gene (Catf1111) and the coded CATP1111  
CC protein  
FH Key Location/Qualifiers  
FT source 1. .2060  
FT location/Qualifiers  
FT 1. .2060 /organism='Candida albicans'.  
FT location/Qualifiers  
FT 1. .2060 /organism='Candida albicans'  
FT /mol\_type='genomic DNA'  
FT /db\_xref='taxon:5476'

FEATURES  
source  
1. .2060  
/organism='Candida albicans'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:5476'

ORIGIN  
Query Match 100.0%; Score 1239; DB 6; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 5e-190;  
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTATCA 60

Db 720 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTATCA 779

QY 61 CGTCCCAAAAAGTATTTTGACATATGAAGGGTGTGATTAAGCCATTAATCGACATCA 120

Db 780 CGTCCCAAAAAGTATTTTGACATATGAAGGGTGTGATTAAGCCATTAATCGACATCA 839

QY 121 TTATTAGCAACATTTAAGAACCCACGATATGATGATCCGATTAATTTGACATGAC 180

Db 840 TTATTAGCAACATTTAAGAACCCACGATATGATGATCCGATTAATTTGACATGAC 899

Qy	181	AAATGTGTAAGA	CA	TTTTTCAGAAA	AA	CA	CA	ATTTTGGAA	AA	CA	CA	ATATGTAT	CA	CA	CTCC	240														
Db	900	GATTTGTATAAG	CA	TTTTTCAGAAA	AA	CA	CA	ATTTTGGAA	AA	CA	CA	ATATGTAT	CA	CA	CTCC	959														
Qy	241	GA	AAAAA	AA	CA	CTTCCATTTG	TC	AGTGTGTGT	GT	TA	AA	GGGGTT	TA	AT	TC	CGACA	CTTG	300												
Db	960	GA	AAAAA	AA	CA	CTTCCATTTG	TC	AGTGTGTGT	GT	TA	AA	GGGGTT	TA	AT	TC	CGACA	CTTG	1019												
Qy	301	AAAAGACA	TG	AAAT	CA	CCCAT	TA	CA	AA	AGTC	AT	TA	AA	TGT	TAC	ATT	GA	AAAT	TGT	CA	GA	360								
Db	1020	AAAAGACA	TG	AAAT	CA	CCCAT	TA	CA	AA	AGTC	AT	TA	AA	TGT	TAC	ATT	GA	AAAT	TGT	CA	GA	1079								
Qy	361	GC	ATTTT	TA	AA	CA	TC	AT	CTTT	TA	AG	AC	AT	CA	T	AT	AT	AT	CT	GT	TC	AT	GA	AAAA	CA	CT	TA	420		
Db	1080	GC	ATTTT	TA	AA	CA	TC	AT	CTTT	TA	AG	AC	AT	CA	T	AT	AT	AT	CT	GT	TC	AT	GA	AAAA	CA	CT	TA	1139		
Qy	421	AC	GT	TA	AA	CA	AT	GT	TA	AA	AA	AGTT	TC	CT	CG	AC	CT	TA	AA	AA	TT	AG	CA	CA	TA	TA	AT	TA	480	
Db	1140	AC	GT	TA	AA	CA	AT	GT	TA	AA	AA	AGTT	TC	CT	CG	AC	CT	TA	AA	AA	TT	AG	CA	CA	TA	TA	AT	TA	1199	
Qy	481	AA	AC	AT	CA	TG	GT	GA	T	CT	CT	GT	TA	TC	AT	CA	AT	GT	GA	TC	AT	CT	GT	TG	TTTT	AAAA	AT	TC	540	
Db	1200	AA	AC	AT	CA	TG	GT	GA	T	CT	CT	GT	TA	TC	AT	CA	AT	GT	GA	TC	AT	CT	GT	TG	TTTT	AAAA	AT	TC	1259	
Qy	541	CA	AA	CT	TG	TC	AG	T	TA	TC	CA	ATTT	CA	T	TA	AA	CA	CT	GC	AT	CC	AAA	CT	TA	AT	GT	CT	CT	600	
Db	1260	CA	AA	CT	TG	TC	AG	T	TA	TC	CA	ATTT	CA	T	TA	AA	CA	CT	GC	AT	CC	AAA	CT	TA	AT	GT	CT	CT	1319	
Qy	601	AA	AT	GT	GT	TA	AG	GT	TG	GT	GG	AAAA	AA	GGTT	AT	CT	TC	CA	AT	GT	TA	AG	TC	AT	660					
Db	1320	AA	AT	GT	GT	TA	AG	GT	TG	GT	GG	AAAA	AA	GGTT	AT	CT	TC	CA	AT	GT	TA	AG	TC	AT	1379					
Qy	661	G	AT	T	C	A	C	A	T	G	A	A	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	720
Db	1380	G	AT	T	C	A	C	A	T	G	A	A	A	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	1439
Qy	721	AAAA	TG	AT	T	A	G	T	T	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	780	
Db	1440	AAAA	TG	AT	T	A	G	T	T	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1499	
Qy	781	TT	AA	AG	AA	CT	GA	AG	T	GA	AAAA	AA	TT	A	GA	AA	CC	T	A	T	TA	AG	T	CA	AG	A	T	TA	840	
Db	1500	TT	AA	AG	AA	CT	GA	AG	T	GA	AAAA	AA	TT	A	GA	AA	CC	T	A	T	TA	AG	T	CA	AG	A	T	TA	1559	
Qy	841	A	A	T	T	G	C	A	T	T	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	900	
Db	1560	A	A	T	T	G	C	A	T	T	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1619	
Qy	901	G	A	T	A	G	T	C	T	A	G																			

RESULT 3			
LOCUS	AX489193		
DEFINITION	Sequence 6493 from Patent WO02053728.	1239 bp	DNA
ACCESSION	AX489193		linear
VERSION	AX489193.1		PAT 16-AUG-2002
KEYWORDS	GI:22323205		
SOURCE			
ORGANISM	Candida albicans		
	Candida albicans		
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
REFERENCE			
AUTHORS	1		
TITLE	Romer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K.L.		
JOURNAL	Gene disruption methodologies for drug target discovery		
	Patent: WO 02053728-A 6493 11-JUL-2002;		
	Elitra Pharmaceuticals, Inc. (US)		
FEATURES	Location/Qualifiers		
source	1..1239		
	/organism="Candida albicans"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:5476"		
ORIGIN			

Query Match	99.7%	Score 1235.8	DB 6	Length 1239
Best Local Similarity	99.8%	Pred. No. 1.9e-189		
Matches 1237	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY 1	ATGAGTAAGAGTACGAAACCAATCGATATCATCTTTAAATCTTCTTCTTCATCA	60		
Db 1	ATGAGTAAGAGTACGAAACCAATCGATATCATCTTTAAATCTTCTTCTTCATCA	60		
QY 61	CGTCCCAAAAAGTATTTTGGCACTATGAAAGGGTGTATTAAGCTTAAATCGACATCA	120		
Db 61	CGTCCCAAAAAGTATTTTGGCACTATGAAAGGGTGTATTAAGCTTAAATCGACATCA	120		
QY 121	TTATTAGGCAACATTTAAGAACCCACAGTAATATGACCCGTAAATATGACAGTGAAC	180		
Db 121	TTATTAGGCAACATTTAAGAACCCACAGTAATATGACCCGTAAATATGACAGTGAAC	180		
QY 181	GATTGTGTAAGCATTTTTCAGAAAAATCACTTTGGAAAACATATTTGATATCATTTCC	240		
Db 181	GATTGTGTAAGCATTTTTCAGAAAAATCACTTTGGAAAACATATTTGATATCATTTCC	240		
QY 241	GAAAAAAAACCATTCATTTGTCAGTGTGTGTGTAAAGGGTTAATTCTCCACACATTG	300		
Db 241	GAAAAAAAACCATTCATTTGTCAGTGTGTGTGTAAAGGGTTAATTCTCCACACATTG	300		
QY 301	AAAAGACATGAATACACCCATACAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAA	360		
Db 301	AAAAGACATGAATACACCCATACAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAA	360		
QY 361	GCATTTTATTAACAATCAATCTTTAAGACATCATATATTTCTGTTCATGAAAAAACATTA	420		
Db 361	GCATTTTATTAACAATCAATCTTTAAGACATCATATATTTCTGTTCATGAAAAAACATTA	420		
QY 421	AAGGTAAACAATTAATAAAGTTTACCTGACCTCAAAATTAGACAAACATTAATTA	480		
Db 421	AAGGTAAACAATTAATAAAGTTTACCTGACCTCAAAATTAGACAAACATTAATTA	480		
QY 481	AAACATCATGTGTGATCTCCGCTTATCAATGTATCATCTGTGTGTTTAAAAAATTC	540		
Db 481	AAACATCATGTGTGATCTCCGCTTATCAATGTATCATCTGTGTGTTTAAAAAATTC	540		
QY 541	CAAACTTGCTAGATTTACAAATTTCAATTAACAACTGACATCCAAAACCTTAAATGTCTT	600		
Db 541	CAAACTTGCTAGATTTACAAATTTCAATTAACAACTGACATCCAAAACCTTAAATGTCTT	600		
QY 601	AAATGTGTAAAGTGTGTGGGAAAAAGGTTATCTTCAACATATGTAAAGCATGAT	660		
Db 601	AAATGTGTAAAGTGTGTGGGAAAAAGGTTATCTTCAACATATGTAAAGCATGAT	660		
QY 720	GATTCTACATGATCAAAATATGACATTTGTGATTTGTGATGTGGGAAATTTGCAAG	720		

Db 661 GATTCTACCATGATCAAAATATGACTTGTGATTATTGTGATGCGGAAATTTGCAAAAG 720  
QY 721 AAAAATGAATTAGTTGAACATTAATAATCTTCATGATGATGTAATATCCCTGATGATTA 780  
Db 721 AAAAATGAATTAGTTGAACATTAATAATCTTCATGATGATGTAATATCCCTGATGATTA 780  
QY 781 TTTAAAGAACTGAAGTGAAGAAATTTAGAACCTTAATTGATCAAGATGCAAAATTTAAAT 840  
Db 781 TTTAAAGAACTGAAGTGAAGAAATTTAGAACCTTAATTGATCAAGATGCAAAATTTAAAT 840  
QY 841 AATTGCAATGAATTTAGAAACAGAGAAATTTAAAGTGAAGAAAGATGAAGAGATGAAGAA 900  
Db 841 AATTGCAATGAATTTAGAAACAGAGAAATTTAAAGTGAAGAGATGAAGAGATGAAGAA 900  
QY 901 GATAGCTAGATGAAAAAGAAAGATGTTAGTTCAGACTCACTCATGCTCAAGATCA 960  
Db 901 GATAGCTAGATGAAAAAGAAAGATGTTAGTTCAGACTCACTCATGCTCAAGATCA 960  
QY 961 ATTTAAATCATTTTCTGCTCTTTGGAAAGTTCAAAGAGTTTCTTAACTTATTTGAAAT 1020  
Db 961 ATTTAAATCATTTTCTGCTCTTTGGAAAGTTCAAAGAGTTTCTTAACTTATTTGAAAT 1020  
QY 1021 AGTGGAGAGATGCAATTTGCTTAAGAAATTAATTGTGATGAAATGTTTCTAGAGAAATAT 1080  
Db 1021 AGTGGAGAGATGCAATTTGCTTAAGAAATTAATTGTGATGAAATGTTTCTAGAGAAATAT 1080  
QY 1081 GATTTCATGTCATGATTTGAATGTCATGATGATTAATAAGAAATTTGATGATCTTA 1140  
Db 1081 GATTTCATGTCATGATTTGAATGTCATGATGATTAATAAGAAATTTGATGATCTTA 1140  
QY 1141 AATGATATGAAAAAGAAAGAACTCCAGAAAGTGAACATTTGTTAAAAAGCCAGAGATG 1200  
Db 1141 AATGATATGAAAAAGAAAGAACTCCAGAAAGTGAACATTTGTTAAAAAGCCAGAGATG 1200  
QY 1201 GATTTCATGTCATGATTTGATGATTTTCTGCAATA 1239  
Db 1201 GATTTCATGTCATGATTTGATGATTTTCTGCAATA 1239

RESULT 4  
AR550053  
LOCUS AR550053 627 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 5184 from patent US 6747137.  
ACCESSION AR550053  
VERSION AR550053.1 GI:53943228  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 627)  
AUTHORS Weinstock,K.G. and Bush,D.  
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics  
JOURNAL Patent: US 6747137-A 5184 08-JUN-2004;  
FEATURES  
source  
1..627  
/organism="Unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 47.2%; Score 584.8; DB 6; Length 627;  
Best Local Similarity 97.4%; Pred. No. 1.3e-84;  
Matches 608; Conservative 0; Mismatches 7; Indels 9; Gaps 1;  
QY 616 TGTGTTGGAAAAAGTTTATCTTCAATATGTTAAGTCATGATGATTTCTACATGATC 675  
Db 13 TGTGTTGGAAAAAGTTTATCTTCAATATGTTAAGTCATGATGATTTCTACATGATC 72  
QY 676 AAAATATGACTTGTGATTTGATGTTGGGAAATTTGCAAGAAAAATGATTAATGTT 735  
Db 73 AAAATATGACTTGTGATTTGATGTTGGGAAATTTGCAAGAAAAATGATTAATGTT 132  
QY 736 GAACATTAATATATCTTCATGATGTTAATATCCCTGATGATTAATTAAGAACTGAA 795

Db 133 GAAATTTTAAATATCTTCATGATGATTAATATCCCTGATGATTTTAAAGAACTGAA 192  
QY 796 GTGAAAAATTTAGAACCTTATAGATCAAGATGCAAAATTTAAATTTGATGAATTA 855  
Db 193 GTGAAAAATTTAGAACCTTATAGATCAAGATGCAAAATTTAAATTTGATGAATTA 252  
QY 856 GAAACAGAAATTTAAAGTGAAGAAAGATGAAGAAATGAAGATAGTCTAGATGA 915  
Db 253 GAAACAGAAATTTAAAGTGAAGAAAGATGAAGAAATGAAGATAGTCTAGATGA 303  
QY 916 AAAAGATGATGTTAGATCAGATCAATGTCAGCTCAAAGATCAATTAATCAATTTACT 975  
Db 304 AGAAGAGATGATTTAGATCAGATCAATGTCAGCTCAAAGATCAATTAATCAATTTACT 363  
QY 976 GCTTCTTTGGAAGTTTCAAAGAGTGTTCCTAACTTAATTTCTGATATGAGGAAAGATC 1035  
Db 364 GCTTCTTTGGAAGTTTCAAAGAGTGTTCCTAACTTAATTTCTGATATGAGGAAAGATC 423  
QY 1036 AATTGCTTAAGAAATATTTGTGATGAATGTTTCTAGAGAAATATGATTTACGTGACAT 1095  
Db 424 AATTGCTTAAGAAATATTTGTGATGAATGTTTCTAGAGAAATATGATTTACGTGACAT 483  
QY 1096 TTTGAATGCGATGATGATTAATTTAACAAGATTTGATCTTTTAAATATGATGAATA 1155  
Db 484 TTTGAATGCGATGATGATTAATTTAACAAGATTTGATCTTTTAAATATGATGAATA 543  
QY 1156 GAAAGAACTCCGAAAGTGAACCATTTGTTAAAAAGCCAGAGATGATTTATGCAAT 1215  
Db 544 GAAAGAACTCCGAAAGTGAACCATTTGTTAAAAAGCCAGAGATGATTTATGCAAT 603  
QY 1216 GAACATGATGATTTCTGATTA 1239  
Db 604 GAACATGATGATTTCTGATTA 627

RESULT 5  
CR382135 06/c  
WPCOMMENT  
Sequence split into 16 fragments LOCUS CR382135 Accession CR382135  
Fragment Name Begin End  
CR382135\_00 1 110000  
CR382135\_01 100001 210000  
CR382135\_02 200001 310000  
CR382135\_03 300001 410000  
CR382135\_04 400001 510000  
CR382135\_05 500001 610000  
CR382135\_06 600001 710000  
CR382135\_07 700001 810000  
CR382135\_08 800001 910000  
CR382135\_09 900001 1010000  
CR382135\_10 1000001 1110000  
CR382135\_11 1100001 1210000  
CR382135\_12 1200001 1310000  
CR382135\_13 1300001 1410000  
CR382135\_14 1400001 1510000  
CR382135\_15 1500001 1592360  
Continuation (7 of 16) of CR382135 from base 600001 (CR382135 Debaryomyces hanseoli chro  
Query Match 35.4%; Score 438.4; DB 8; Length 110000;  
Best Local Similarity 63.4%; Pred. No. 1.4e-61;  
Matches 724; Conservative 0; Mismatches 406; Indels 12; Gaps 3;  
QY 33 ATCTTAATATCTTCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 92  
Db 20471 ATCTTAATATCTTCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 20412  
QY 93 GTGATTAAGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 152  
Db 20411 CTGCGGAAAGCCCTTATTTTAAACATTCCTTGAACAACATTAATAAGATCAATAC 20352  
QY 153 TGATGACCGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 212



```
Db 20351 TGAAGACCTTTAAATGCTCGGAACTGGATCGCAAAATCATCTCTTAAGAAAGTCACA 20292
Qy 213 TTTGGAACACATATGTATC-----ACATTCGGAAAAAACCATTCCATTGTTCAGT 266
Db 20291 TTTACAGACCATTTCTTCTCATGAAGATCAAGAAAGTAAACCTTCCAGGTGCAAC 20232
Qy 267 GTGTGGTAAAGGGTTAAATCTCGACACACTTGAAGAGCATGAATCAACCCATCAAA 326
Db 20231 TTGGCGCAAAAGGTGTAATCCCTTACACACTTTAAAGAGACAGAAATTAACATACAA 20172
Qy 327 GTCAATTAATGTACATTTGAAAAATGTCAAGAGCATTTTATTAACATCAATCTTTAAG 386
Db 20171 ATCAATTCATTGTACATTTGAGGGCTGTATGATCATTTTATTAACATCAAGTCACTG 20112
Qy 387 ACATCATATATTTCTGTTTCATGAAAAACATTAAGGTGAACATATGTAATAAGTTT 446
Db 20111 ACACCATACGCTATCGGTTTCATGAAAAAATTAATTAAGCAATGTAATTAATCGTT 20052
Qy 447 CACTCGACCTTCAAAATTAAGCAACATTAATTAACATCATGTGTGATCTCTGCTTA 506
Db 20051 TTAACAGACCGTATCGATTGGACACAGCAATTAATTAATCAAGTGAATCGCCGCTTA 19992
Qy 507 TCAATGTATCATCTGTTGTTTAAATTTCAAACTTGGTCAGTATTACAATTTCA 566
Db 19991 TCAATGTATCATCAAGGCTGTTTGGAAATTTATGAGGTGCGGCTTTACAATTTGCA 19932
Qy 567 TATTAACAACATGCAATCCAAACTTAATGTCTTAATGTGTAAAGTTGTGTGGAA 626
Db 19931 TATTAACAACATGCAACCCAAAGTAAAGTCTTAATGTGTGGAAAGTTGCTGTGGTAA 19872
Qy 627 AAAAGTTATCTTCATATGTTAAGTCAATGATTCATCATGATCAAAATATGAC 686
Db 19871 GAAAGGCTTGGTCAACATGATTTTCAAGTGAAGAAAGATGTCAATTTATGAA 19812
Qy 687 TTGTGATTAATGTATGTGGGAAATTTGCAAGAAATTAATGTTGAACATTATTA 746
Db 19811 TTGTATTAATGTATGATTTGAAATTTTGAAGAAAGCTGATTTGATGATCATTAACA 19752
Qy 747 TATCTTCATGATGTATATTCCTGTATGATTTTAAAGAACTGAAAGTAAATTT 806
Db 19751 CACTTATCAAGACAAATCTTCTGTATGATTTTAAACCGATCGAAGCAAGACGTT 19692
Qy 807 AGAGAACCTATTA--GATCAAGATCGAAATTAATTAATTTGATGAATTAAGAACAG 863
Db 19691 AGATTAACCTTAAGGAGAGGACTCTACCAATTTGATTTTAATCTTTAGAGGATTT 19632
Qy 864 GAAATTAAGGTGAAGAGATGAAGAGATGAAGAGATGATCTAG--ATGAAAAAG 920
Db 19631 ACAATCAAAAGGTTTGTGAAGTACCGTGGAGTGAAGAGAGAGAGATGATTAAT 19572
Qy 921 AAGTATGTTAGTCAACTCAATGTCAAGTCAAGATCAATTAATCATTTAATCTGCTTC 980
Db 19571 GTCCATATGTTTCAAGCTTCAATTAATCTTTGAATTCGTTGAATTTAATCTTTGAATC 19512
Qy 981 TTTGGAAGGTCAAGAGTCTTCTAACTTATTCGAATGATGGAAGAGATCAATTC 1040
Db 19511 TGGAAAGCATCTATGTATGATCTTATCTGAATTAATTAATCTGAAAAAGAAATTCATG 19452
Qy 1041 TCTTAAGATTAATTTGATGATGATGTTTCTAGAGATATGATTTAGTGAATTTGAA 1100
Db 19451 TCCGAAAAAGATTTGATGATGATGTTTCAAGCGGTGATTAATCTTGAAGAGACATTTAA 19392
Qy 1101 ATGGCATGATGATTAATTTCAAGAGATGATGATCTTAAATAGTATAGAAAAAGAGA 1160
Db 19391 GTGGCATGATCTCACTTAAGAAAGATTAAGAGGTTTAAAAAGCTTGAAGTGAAGA 19332
Qy 1161 AA 1162
Db 19331 AA 19330
```

RESULT 6  
AR550055

```
LOCUS AR550055 462 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 5186 from patent US 6747137.
ACCESSION AR550055
VERSION AR550055.1 GI:53943230
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 462)
AUTHORS Weinstock,K.G. and Bush,D.
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics
JOURNAL Patent: US 6747137-A 5186 08-JUN-2004;
FEATURES
source location/Qualifiers
1..462
/organism="unknown"
/mol_type="genomic DNA"
```

```
ORIGIN
Query Match 31.6%; Score 391; DB 6; Length 462;
Best Local Similarity 98.5%; Pred. No. 2.1e-53;
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
```

```
Qy 1 ATGAGTGAAGTGAACGAAACCAATGATATCATCTTATATCTTCTTTCATCA 60
Db 52 ATGAGTGAAGTGAACGAAACCAATGATATCATCTTATATCTTCTTTCATCA 111
Qy 61 CGTCCCAAAAGTATTTTGGACATATGAAGGGTGTATTAAGCTTATTAATGACATCA 120
Db 112 CGTCCCAAAAGTATTTTGTATATGAAAGGGTGTATTAAGCTTATTAATGACATCA 171
Qy 121 TTATTGACCAATTTAAGACCCACAGTAATGATCGACCGTAAATGATAGTGAAG 180
Db 172 TTATTGACCAATTTAAGACCCACAGTAATGATCGACCGTAAATGATAGTGAAG 231
Qy 181 GATTGATTAAGATTTTTCAGAAATGACATTTGGAACACATATTTGATCAATTC 240
Db 232 GATTGATTAAGATTTTTCAGAAATGACATTTGGAACACATATTTGATCAATTC 291
Qy 241 GAAAAAACAATTCATTTGTCAGTGTGTGAAGGGGTAAATCTGCAACAACCTTG 300
Db 292 GAAAAAACAATTCATTTGTCAGTGTGTGAAGGGGTAAATCTGCAACAACCTTG 351
Qy 301 AAAAGCATGAATTCACCATACAAAGTCAATTTAATGTATCAATTTGTAAGAA 360
Db 352 AAAAGCATGAATTCACCATACAAAGTCAATTTAATGTATCAATTTGTAAGAA 411
Qy 361 GCATTTAT-AAACATCAATCTTAAAGATCATATATTTATCTGTCTATGA 410
Db 412 ACATTTATTAACATCAATCTTAAAGATCATATATTTATCTGTCTATGA 462
```

```
RESULT 7
CR382122 03/c
WPCOMMENT
Sequence split into 14 fragments LOCUS CR382122 Accession CR382122
```

Fragment Name	Begin	End
CR382122_00	1	110000
CR382122_01	100001	210000
CR382122_02	200001	310000
CR382122_03	300001	410000
CR382122_04	400001	510000
CR382122_05	500001	610000
CR382122_06	600001	710000
CR382122_07	700001	810000
CR382122_08	800001	910000
CR382122_09	900001	1010000
CR382122_10	1000001	1110000
CR382122_11	1100001	1210000
CR382122_12	1200001	1310000
CR382122_13	1300001	1320834

Continuation 74 of 14) of CR382122 from base 300001 (CR382122 Kluyveromyces fragilis str)

Query Match 18.4%; Score 228; DB 8; Length 110000;  
 Best Local Similarity 58.1%; Pred. No. 9.8e-28;  
 Matches 443; Conservative 0; Mismatches 310; Indels 9; Gaps 2;

41 TATCTTCTTCTTCTTATCATGCGCCAAAAGATATTTTGACATATGAAAGGTGTGATA 100  
 15150 TGTCACTTTCATGTCACAGCGCTTAAAGTATATTTCTGCAATTAATCAAGGGTGTAA 15091  
 101 AAGCCTAATATGAGCCATGATTAATAGCAATTTAAGAAACCAAGTAAATGATCGAC 160  
 15090 AATCGTTCAAGACCGAGTTGGTCTCAGGAGATCAGCTCAC--CGATCATAGGAA 15034  
 161 CGTAAATATGACAGTGAAGCATTTGATTAAGCATTTTTCAGAAATCAATTGGAAA 220  
 15033 TCMAACCGTTTAAATGCGACACTTGTGGCAAAAGAAATTTGCAAAAATGCAATCTGAATA 14974  
 221 CACATTTGTATCACTTCGCAAAAAACATTCATGTTTCAAGTGTGTGTAAGGGG 280  
 14973 GACACATGTTCTCTCTCATGATGATTAACCTTTCACATGTTCTATATGTGGAAAGGTG 14914  
 281 TTAAATTCGACACACATTTGAAAAGACATGAATTCACCCATCAAAAGTATTAATGTA 340  
 14913 TCACACACGACACAGATTAAAGACATGAATTAACATAGCAAGTCTTCCACTGT 14854  
 341 CATTTGAAATTTGTCAAGACATTTTAAACATTAATCTTTAAGACATCATATATAT 400  
 14853 CATATGAGGAGATGTAACGATATCTTCAACAACTCCACACTCCATCGCATATCTGT 14794  
 401 CGTTTCATGAAAAACATTAAGGTGAACATGTAATTAAGTTTTCATCGACCTTCA 460  
 14793 CGTTTCAGAAAAGAACTTACCTGCTCCTGACTGTAACAAAGCTTCCAAAGGCTTATC 14734  
 461 AATTAGACACATTAATTAATAACATCATGTGATCTCTGC-----TTATCATATG 514  
 14733 GATTAAGATACATATGATTAAGCATCAACACCGAATCTACCGAATGTACATGTG 14674  
 515 ATCATCTGTGTTTAAAAATTTCCAACTTGTGATGATTAATTAATTAATTAATTAAC 574  
 14673 ACTTCTTTCATGATCCGACGATTTTCCACTGTGCTTCTTACCAACATATTAATTAAC 14614  
 575 AACTGATCCAAAACCTTAATGTCCTAATGTGTAAAGTGTGTGGGAAAAAGTT 634  
 14613 AATGCCATCGAAATTAACATGATCTTATTTGGGAAACCTTGTGTGATCGGT 14554  
 635 TATCTTCAATATGTTAATGATGATGATTTCAACATGATCAAAATATGACCTTGAT 694  
 14553 TACGGAATCATATGATGATTTCTATGACGAGATTTGGTCTCTAAGATTTGGAATGTCTT 14434  
 695 ATTTGATGTGGGAAATTTGCAAGAAATGAATTAAGTGAACATTAATATCTTCC 754  
 14493 CATGTCAGATATCATCATTTGGAAGAAATCTCAGCTATACAACTATTAAGAAAGTC 14434  
 755 ATGATGCTAATATCCCTGATGATTTATTAAGAAATCTAAG 796  
 14433 ACMAAGATGTAATCCAGAAATGCTATTAATCTGCCGACGACG 14392

RESULT 8  
 CR380953 09/c  
 WPCOMMENT  
 Sequence split into 10 fragments LOCUS CR380953 Accession CR380953

Fragment Name	Begin	End
CR380953_00	1	110000
CR380953_01	100001	210000
CR380953_02	200001	310000
CR380953_03	300001	410000
CR380953_04	400001	510000
CR380953_05	500001	610000
CR380953_06	600001	710000
CR380953_07	700001	810000
CR380953_08	800001	910000
CR380953_09	900001	992211

Continuation (10 of 10) of CR380953 from base 900001 (CR380953 Candida glabrata strain C

Query Match 18.3%; Score 227.2; DB 8; Length 92211;  
 Best Local Similarity 57.8%; Pred. No. 1.4e-27;  
 Matches 447; Conservative 0; Mismatches 318; Indels 9; Gaps 2;

44 CTCTTCTTCTTCTTATCATGCGCCAAAAGATATTTTGACATATGAAAGGTGTGATA 103  
 60262 CATCTGATATCCAGTGAAGGTGGAAGACATCTTTTGAAATACATATATTTGTCCAAAG 60203  
 104 CCTATATCGACCATCATTTATTAAGCAACATTTAAGAAACCAAGTAAATGATCGACGT 163  
 60202 CCTTTAAAGGCGCTCTTCTGCTACAGAACACCAAGATATC--AATGATCTTGGAGGA 60146  
 164 ATAAATGTAAGTGAAGCATTTGTGATTAAGCAATTTTTCAGAAATCAATTTGGAACAC 223  
 60145 AACCTGGAATATGTAATCAATGTAAGATTCATTAACAAAAGATTCACCTTAGAGAGAC 60086  
 224 ATATTGTATCATCTTCGAAAAAAACCATTCATTTGTCTGCTGTGTGTAAGGGCTTA 283  
 60085 ACTTATACACACACAGATGAAGACCGTTTATTTGTTCAATTTTGTGGAAAGGGCTTA 60026  
 284 ATTCGACACACATTTGAAAAGACATGAATCACCCATACAAAGTATTAATGATCAT 343  
 60025 TTACTAGGCAACATGTAAGACATGAGTACTCAACCAATCGTTCAATTTGTGANT 59966  
 344 TTGAAATTTGTCAAGACATTTTAAACATCAATCTTTAAGACATATATATCTG 403  
 59965 ATGAAGTTGTATGAGAGTTTCTACAGACCTTCATTAAGGGCTCATATTTTAGCAG 59906  
 404 TTCTGAAAAAACAATTAACGTGTAAACATGTAATTAAGTTTCTGACCTTCAAAAT 463  
 59905 TTCAATTTCAAAAGTCTAAATATGTCAGATGCAACMAAGTTTTCMAAGACCTTAACAGAC 59846  
 464 TAGCACAACATTAATTAACATCAT-----GGTGAATCTCCGCTTATCATGATGATC 517  
 59845 TCAAAAATCAATAGCCAAACCAATATCCAGATGTTGTCAATCTTAATATATACCT 59786  
 518 ATCTGTTGTTTAAAAATTTCCAACTTGTGATGATTAATTAATTAATTAATTAAC 577  
 59785 TCAGTGTGTTGTTCTAAGAGTTTCAAAACATGATGCTGCTTAAGATTAATTAATTAAG 59726  
 578 TGCATCCAAAACCTTAATATGTCCTAATATGCTGAAGTTGTGTGGAAAAAGTTTAT 637  
 59725 ATCATCCGAAGTTAAATAGCCCTATATGACATTAACCTTGTGTGGGAAAGCGTTTAA 59666  
 638 CTTCATATGTTAAGTATGATGATGATTTACATGATCAAAATATGACCTGTGATAT 697  
 59665 ATATGACATGAAATCCAGATGAACCTTGTTCAGAAATTTGAAGTGTCTATAT 59606  
 698 GTGATGTGGGAAATTTGCAAGAAATGAATTAAGTGAACATTAATTAATTTCCATG 757  
 59605 GTAATGATCATGATTTCTTAAGAAATTAAGAACTTTGATCATATACGCAACATCATTA 59546  
 758 ATGTAATATCCCTGATGATTTATTAAGAAACTGAAGTGAAGAAATTAAGAGA 811  
 59545 GTGAAGAAATCCGCTATATTAAGAAAGAGTGTGCTTATGTTGAAA 59492

RESULT 9  
 YSCNOFEAT  
 LOCUS YSCNOFEAT 1560 bp DNA linear PLN 27-APR-1993  
 DEFINITION Saccharomyces cerevisiae transcription factor IIIA and RNA polymerase subunit RPB6 genes.  
 ACCESSION M90638.1 GI:1172046  
 VERSION M90638.1  
 KEYWORDS RNA polymerase; transcription factor IIIA.  
 SOURCE Saccharomyces cerevisiae (baker's yeast)  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 REFERENCE 1 (bases 1 to 1560)  
 Woychik, N.A. and Young, R.A.  
 Genes encoding transcription factor IIIA and the RNA polymerase

common subunit RPB6 are divergently transcribed in *Saccharomyces cerevisiae*  
Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)

JOURNAL  
MEDLINE  
PUBMED

92237295  
1570925

COMMENT Original source text: *Saccharomyces cerevisiae* (strain S288C) DNA.

FEATURES

source  
1. .1560  
Location/Qualifiers  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="S288C"  
/db\_xref="taxon:4932"

ORIGIN chromosome 14.

Query Match 17.5%; Score 217.4; DB 8; Length 1560;  
Best Local Similarity 59.4%; Pred. No. 1.4e-25;

Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

QY 62 GTCCCAAAAGTATTGTCATATGAAAGGTGTGTAAGGCTTAATTCAGACATCAT 121  
DB 368 GGCCAAGACATATTTCTGTGATGATGATGGCTGTGTAAGGATTTACAAGACCTTCAA 427  
QY 122 TATTAGAGCAACATTTAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGACG 181  
DB 428 TTTTGACTGAAACCAATTAAGCGTACATCAGGGTTTA--AGAGCATTTCAGTGTGATA 484  
QY 182 ATTGTGATTAAGATTTTTCAGAAATTCATATTGGAACACATTTGATTCATTCGCG 241  
DB 485 AGGTGCAAAATCTTGTTAAAGAGTACTTAAAGACACTGTGATACGATTCG 544  
QY 242 AAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACACATTCGA 301  
DB 545 ATACGAAACATTCATTTGTCATTTATTTGTAAGAAAGAGTGAAGACTCCGACACATCGA 604  
QY 302 AAAGACATGAATACCCCATACAAAGTCAATTTAATGTACATTTGAAAAATTGTCAAAG 361  
DB 605 AGGACACAGAAAGTACGATACCAAACTTTCAATTTGTCCAGAGAAAGATGCAACCTCC 664  
QY 362 CATTTATTAACATCATCTTTAAGACATATATATATCTGTTCATGAAAAACATTTAA 421  
DB 665 GATTTCTACAAAGCATCCCAATTAAGGACATATTTATCTGTTCATTTACATTAACCTAA 724  
QY 422 CGTGAACAAATGTAATTAAGTTTTCACGACCTTCAAAATTAAGCAACATTAATTA 481  
DB 725 CCGTCCACACCTGCAATTAAGGCTTTCAGAGGCCATTAAGGCAATTAACATTTCTTA 784  
QY 482 AACATCA-----TGGTGAATCTCGCTTATCAATGTGATCAATCTGTTGTTTAA 535  
DB 785 AACATCAAGATCTGAGGTGAAATAATCTTAACCAATGTACTTTTGTGTTGTAAG 844  
QY 536 ATTTCCAACTTGGTCAATTTAATTTAATTAACCACTGACATCCAAAATTAAAT 595  
DB 845 AGTTTGGATATGATGTCATATTCGATTAATAAATTAATGATCAATCCAAAATTAAT 904  
QY 596 GTCCCTAAATGTGAAAGGTGTGTGGGAAAAAAGGTTTATCTTCACATATTTAAGTC 655  
DB 905 GTCTATTTTATAGCAACCATGTGTGGGAAAAATGTTTTCAAAATGACATATTAATTC 964  
QY 656 ATGATGATTTACATGATCAAAATATGACCTTGTGATTAATGTGTGATGTGGGAAAAATT 715  
DB 965 ATACGACCTCAGTAAAGTGAAGGTGATATATATGTCCGATATATGCTTTT 1024  
QY 716 CAAGAAGAAATGAATTAAGTTGAACATTTAT 744  
DB 1025 CTAGAAAAACATGATCTTCTCAGCATTTAT 1053

RESULT 10  
YSCFTTIIIA 1739 bp DNA linear PLN 13-SEP-1996  
LOCUS *Saccharomyces cerevisiae* transcription factor IIAA (TFIIIA) gene,  
DEFINITION complete cds.  
ACCESSION M80611

VERSION M80611.1 GI:172902  
KEYWORDS transcription factor IIAA; zinc-finger protein; zinc-finger transcription factor.

SOURCE *Saccharomyces cerevisiae* (baker's yeast)

ORGANISM *Saccharomyces cerevisiae*

REFERENCE 1 (bases 1 to 1739)  
*Saccharomyces cerevisiae*, *Saccharomycetaceae*; *Saccharomyces*.

AUTHORS Archambault, J., Milne, C.A., Schapert, K.T., Baum, B., Friese, J.D.

TITLE The deduced sequence of the transcription factor TFIIIA from *Saccharomyces cerevisiae* reveals extensive divergence from *Xenopus*

TFIIIA

JOURNAL J. Biol. Chem. 267 (5), 3282-3288 (1992)

MEDLINE 92147684

PUBMED 1737784

FEATURES

source

1. .1739  
Location/Qualifiers  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="W303-1A"  
/db\_xref="taxon:4932"  
/chromosome="XVI"  
/gene="TFIIIA"  
234. .1523  
234. .1523  
/gene="TFIIIA"  
/note="zinc finger containing protein"  
/codon\_start=1  
/product="transcription factor IIAA"  
/protein\_id="AAB08014.1"  
/db\_xref="GI:172903"

gene

CDS

62 GTCCCAAAAGTATTGTCATATGAAAGGTGTGTAAGGCTTAATTCAGACATCAT 121  
DB 367 GGCCAAGACATATTTCTGTGATGATGATGGCTGTGTAAGGATTTACAAGACCTTCAA 426  
QY 122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGACG 181  
DB 427 TTTTGACTGAAACCAATTAAGCGTACATCAGGGTTTA--AGAGCATTTCAGTGTGATA 483  
QY 182 ATTGTGATTAAGATTTTTCAGAAATTCATATTGGAACACATTTGATTCATTCGCG 241  
DB 484 AGGTGCAAAATCTTGTTAAAGAGTACTTAAAGACACTGTGATATGACATTTCTTA 543  
QY 242 AAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACACATTCGA 301  
DB 544 ATACGAAACATTCATTTGTCATTTATTTGTAAGAAAGAGTGAAGACTCCGACACATCGA 603  
QY 302 AAAGACATGAATACCCCATACAAAGTCAATTTAATGTACATTTGAAAAATTGTCAAAG 361  
DB 604 AGGACACAGAAAGTACGATACCAAACTTTCAATTTGTCCAGAGAAAGATGCAACCTCC 663  
QY 362 CATTTATTAACATCATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTTAA 421  
DB 664 GATTTCTACAAAGCATCCCAATTAAGGACATATTTATCTGTTCATTTACATTAACCTAA 723  
QY 422 CGTGAACAAATGTAATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 481  
DB 724 CCGTCCACACCTGCAATTAAGGCTTTCAGAGGCCATTAAGGCTTAAGAAATTAACATTTCTTA 783

ORIGIN

Query Match 17.5%; Score 217.4; DB 8; Length 1739;  
Best Local Similarity 59.4%; Pred. No. 1.3e-25;

Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

QY 62 GTCCCAAAAGTATTGTCATATGAAAGGTGTGTAAGGCTTAATTCAGACATCAT 121  
DB 367 GGCCAAGACATATTTCTGTGATGATGATGGCTGTGTAAGGATTTACAAGACCTTCAA 426  
QY 122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGACG 181  
DB 427 TTTTGACTGAAACCAATTAAGCGTACATCAGGGTTTA--AGAGCATTTCAGTGTGATA 483  
QY 182 ATTGTGATTAAGATTTTTCAGAAATTCATATTGGAACACATTTGATTCATTCGCG 241  
DB 484 AGGTGCAAAATCTTGTTAAAGAGTACTTAAAGACACTGTGATATGACATTTCTTA 543  
QY 242 AAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACACATTCGA 301  
DB 544 ATACGAAACATTCATTTGTCATTTATTTGTAAGAAAGAGTGAAGACTCCGACACATCGA 603  
QY 302 AAAGACATGAATACCCCATACAAAGTCAATTTAATGTACATTTGAAAAATTGTCAAAG 361  
DB 604 AGGACACAGAAAGTACGATACCAAACTTTCAATTTGTCCAGAGAAAGATGCAACCTCC 663  
QY 362 CATTTATTAACATCATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTTAA 421  
DB 664 GATTTCTACAAAGCATCCCAATTAAGGACATATTTATCTGTTCATTTACATTAACCTAA 723  
QY 422 CGTGAACAAATGTAATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 481  
DB 724 CCGTCCACACCTGCAATTAAGGCTTTCAGAGGCCATTAAGGCTTAAGAAATTAACATTTCTTA 783

Oy	482	AACATCA-----TGTGATCTCTGGCTTACATGATCATCTGGTGTTTAAA	535
Db	784	AACATCAGATCTGAGTAGAATACTCTTACCAATGACTTTGTGTGTTGCAAG	843
Oy	536	ATTTCAAACTTGTGCAGTATTCAATTCATATAAAACAATGCATCCAAACTTAAAT	595
Db	844	AGTTTCGATATGTGCACAATGCATATGCAATATAAATAATGATCATCTTAATTTAAAT	903
Oy	596	GTCCTAAATGTGTAAAGTTGTGTGGAAAAAAGTTTATCTTCACTATAGTTAAGTC	655
Db	904	GTTCTATTGTGCAACCATGTGTGGGGAAAAATGGTTTACAAATGCACATGATTATTC	963
Oy	656	ATGATGATTTACATCATGATCAAAATATGACTTGTGATTAATTGTGATGTGGGAATTTTG	715
Db	964	ATGACGACTCACTAGTAACCAAAAATGTGAAGTGCATATATATGTCCTGATATGCTTTTT	1022
Oy	716	CAAGAAAAATGAATTAAGTTGAACCTTAT	744
Db	1024	CTAGAAAAATGATCTTCTCAGCACTTAT	1052

RESULT 11					
YSCP677/c					
LOCUS	YSCP677	37497 bp	DNA	linear	PLN 01-AUG-1997
DEFINITION	Saccharomyces cerevisiae chromosome XVI	cosmid 9677.			
ACCESSION	U25841 U00094				
VERSION	U25841.1 GI:786295				

**SOURCE** *Saccharomyces cerevisiae* (baker's yeast)  
**ORGANISM** *Saccharomyces cerevisiae*  
*Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.*

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 37497)	Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Du, Z., Favetto, A., Fulton, L., Gattung, S., Greco, T., Kirszen, J., Kucaba, T., Hallsworth, K., Hawkins, J., Hillier, L., Jier, M., Johnson, D., Johnson, L., Langston, Y., Latreille, P., Le, T., Mardis, E., Meneses, S., Miller, N., Nhan, M., Pauley, A., Peisno, D., Rikken, L., Riles, J., Tatch, A., Trevaastri, E., Vignati, D., Wilcox, L., Woldman, P., Vaudin, M., Wilson, R. and Waterston, R.	The sequence of a portion of the right arm of Saccharomyces

JOURNAL Unpublished (1995)  
REFERENCE 2 (bases 1 to 37497)

TITLE	REFERENCE
The sequence of <i>S. cerevisiae</i> cosmid 9677 Unpublished (1995)	3 (bases 1 to 37497)

**JOURNAL COMMENT** Submitted (27-APR-1995) Robert Waterston  
Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA  
e-mail: mj@sequencer.wustl.edu

NEIGHBORING COSMID INFORMATION:  
This sequence includes nucleotides 1-37497 of cosmid 9677. The  
cosmid on the left is 9705.

FEATURES	Location/Qualifiers
source	1. .37497

```
/organism="Saccharomyces cerevisiae"  
/mol type="genomic DNA"
```

```

/strain="S288C (AB972)"
/db xref="taxon:4932"

```

```

/maþ="XVI"
/notes="baker's vaant"

```

gene complement (2366, .365)  
/gene~"ITFC2"

complement (2366 . .3655)	7 genes: <i>ARC2</i>
CDS	1 gene: <i>ITPA2</i>

2017年11月

```

/note="P9677.9"
/codon_start=1
/product="Transcription factor TFIIA (PIR accession
number S20050)"
/protein_id="AAB66415.1"
/db_xref="GI:786305"
/translation="WGGEVLNNECMPLAELKQETTPISRBSESLSLSTYTRSSSN
RPKYFCYDGDGDKAFTRPSILTEHQLSHVGLAFQCDCAKSPVKSHLIERLYLV
SDTPKPSGCGKGVYTRQDLKRHVVHTKSPICBEGCNILRFYKPOLSHLILSVL
HLLTLPNCKSPFSPYRLRNHISKHGDVEENPVQCTPAGCCGCFRIMSLQSIHNC
HPLKLCPICSKSPGVENGILQWMTIHDSDLVTKMKMCHICPMSPFSKHLITGYCI
HTEEDIPLEIKYKISDIQQLVDHGVGLGNSKHNEDDEKTSNRLEKRLTFENNIV
PLEQVNEVLERLSEGGNGILNLNTVGRKRCYNNCSKPTFKKEYKGIIDKXVH
ELKTLIQLKCKENNTLVQDNHKEPFIILQKETSGDGK"
3890. 4433
/gene="RPB6"
/gene="RPB6"
/join(3890..3909,3986..4433)
/gene="RPB6"
/note="P9677.8"
/codon_start=1

```

```

/protein_id="AA064616.1"
/db_xref="gi:786304"
/translation="MSDYEAFLNDNGNPNFEDVDFHFSDEYVEEKQPFQDGETTDN
KGTIVTGNGSPEDQGHQIQRRTLKEKAIKPDQATTPMTKRYRARIIGTALQIS
MAAPFVLDLSEGTDLRLAMKELAEKTIPLVIRYRLPDGSGFEDMSVEELLVDL"
complement(4621..5112)

```

CDS	complement(4621..5112) /gene="P9677.10" /note="Similar to calmodulin (e.g., Swiss Prot. accession number P13668) and calmodulin-related proteins (e.g.,

```

Swiss Prot. accession number P27482) from several
organism."
/codon_start=1
/evidence=not experimental
/protein_id="AAB661Y.1"
/db_xref="GI:786306"
/transcript="MHSESJLTNQJLTDYIINKIKDAFQMLDBEDBDGLISRGDTIKY
ALGKCHLDDEMSKVPDNDSTIAVEEGVSPFIPLSIINKSLDSPEEELIEESLIK
AIGRCHDLIDNEVLNVDLSKEAGEENEEBEPAKFLKPLTNQQAOTBERTRGKFLPDS

```

gene 5297. .9595  
/gene="SKI3"

```

/ gene="SKI3"
/ note="P9677.7"
/ codon_start=1

```

```

/db_xref="GI:786303"
/translation="MSDIKOLLKEAQELTNRDYEETIEISEKVLKDDPNYAHIFL

```

GKALSLPLASNNVSSNNRLEBATHYHSAKVLVPDNLAMKGFLELRTTBVDPILS  
 YDEFDLOQVADALIKQDSQVELINDIKLKLKTHKADSPYOHKPKSPSLAETIGS  
 RHLSPDQALNLIKLISNFTTEIGLELQNRNLKTHKADSPDYOHKPKSPSLAETIGS  
 ELDQLYNVLNLLDDQKRSIEINQWLEYRLKALKSNPLDVKDFPKFVXEMVEDWVL

KSLESANKPPEGHKTEKETDIDKDVETNEDEYKDRVEDEYKDRVEDEVKQOBEAKE  
DEEDDLDDIEIGLLEEVTVLTENIVCKNNILAHILCOYLLTKEYBAALPIYIKN  
GISLAIYNIKDLGVHLPITKREPSLDLTVYTVVADAPKHNAALKYDNLISGDFSNL

QAKMGKGIIFIERKNWKDAMTLLTOVHESPNNLEVLSELSWSKAHMGYDEBALAGLD  
TVIKGIGKMDLRSIDFRALNLMROAKVIMKHASINDACOENVKCAFKLIIOSIKLD

TPAPGFSTLGDYCHYYKDLRAFCYKPAFDLDAGDYAAKYTTETVASKPNWQAAAS  
SIASRLIGKEKAKAELRSNNWPERVYGAHLEKOEESDSIEMFOSALRYDPNDVESMV

GLGQAYHACGRIEASIKVPDKAIQLRPSHTPAQYFKALSLCDVGEYLSLIDLEKVCQ  
EATRESEFOIGIWEVIMRCSI.DI.YSOGSPLKSVSIAKCTTERIKITIIISEI.KCENQOWA

IYL5QVLRLEFIWIESKVDILPVESTVSIENSQFSGSEELSDVNIKIDILLDSTDD  
 MSTACKELIASKSVSDOKETDIACETBASVYMNITGISELTARTITKEPOYBDAAT

PAFKSIOLOSNTSEWIGLGIATMDINFRVSOHCIFIKATALEPKATNTNENLAMLGL  
WYDEEPAOOUT NKT OSI ABONSDPFCMLT T EEOGDTTCESTI PAUGCTI ONOGR

AAQFMVAKVLENNHNGDDEDIETVEKCTTASIALTEBFKKSPPDSQPALCALTL  
NDITRYNNAVAVI AND TACTI DIVERZEMOCHU ENVAETFOCKASPTQACUMT CTS

ERUHI IENANE LLAN KLUIGI WENI ENI QUBERLE NFALLINGU FAKI INUGU NF EUSI B

gene  
CDS  
MNDLSQGIISSESDKSMKTKISNHIICLGSLYFELNDFDQTLNQFOELISISKSKHL  
VVLIAKVLVDGSDTEKIALOELTEYIATSGADLIATVTLIAAMSILDDKREDLSIIL  
BELKALPLSKOIIDKHKDAPLYIEETITKRLYRNDGKQVORSAYFPFNNKVMYEBLD  
KNIQRIASNGKNTAEMSKICYCSKRLRSIQRMFLCPMNTVAVKALNECF"  
/gene="RPC3"  
/complement (9710. .11674)  
/note="RPC3"  
/note="p9677.11"  
/codon\_start=1  
/product="74 kD subunit of RNA Polymerase III (Swiss Prot.  
accession number P32349)"  
/protein\_id="AAB64619.1"  
/db\_xref="GI:786307"  
/translation="MODEL:GALISAENONGESVSESEKLVTPEDVNTTSLISRLQTLNP  
DLFLYKELVKAHLGERASVITGMVALGSLVRELYEIKDGDVSDVSKTTLVSLQLR  
CVKYLQELTISGKRTYYINEGHIILYSLGIIIDEITQKMNVEEERKQDLVAIV  
QNVISLGSITVEDYLSVSDSMKRTISLFPQLCEMGYLIQISKLATPIEDLQPL  
YEHGYKNI PRNSPLSDIKKRSQAKNAKTDPAKI INKPELSQILITVDPKTSIRIVK  
TVSLTINLRPMKGRSKOLINAKTRVSVTAQVYKIALRLTEOKSPKIRIDPLQGT  
LIQDLLEAKSFQDEALVEKTPGLTPNALDIARHI.PABLDL.RGSLISRKPSNRKS  
GSAALASLPSKLTEDGFTVPALPAVSKSLQSDDTQDEDEEEDLDADTDPHSA  
SLINSLKLIASSNFPFLNETKPGVYVYPSKLPVLSVYEVYIASTLUGPAMLS  
RCTIRDKLVSEKINSTALMKEKDIRSTIASLIRNSVIEQVPRPADSASRAVFLF  
RCKETSYNFMRONLEMMNANILFKKEKTKQENSTLLKKANRDVKGRENELLPSSEL  
NOLKQVNERELNPARLSRLSLMEVFQMA"  
12014. .13120  
/gene="QCR2"  
12014. .13120  
/gene="QCR2"  
/note="p9677.6"  
/codon\_start=1  
/product="ubiquinol-cytochrome c reductase core protein 2"  
/protein\_id="AAB64620.1"  
/db\_xref="GI:786302"  
/translation="MLSAARLQPAQGSVERLTVSARDAPTKISTLAVKHGSGRYATK  
DGAHLNRPNQNTWTRSAIKLVRESELGTFKSTDLREYITLTKATFLKQDLPIYV  
NALADLYKTAPEKHELTSSVLPAAVYDAVAECQVKAEDOLVATIRKIGENIL  
YDCEVYSLODIDKDPADKTYTKENLEVSQENVAEDLKRVEDLSLSTIPAGSLYSK  
SEKPFPLGSENRVRFIDSVAAIGIPNNKASLAQVLANYLTSLSGSLISSKL  
DKETDGLFTLPRDQDSAVSNIKKIVASLIDKQDLSPAINVTKLKNAVONESVSS  
PIELNPDVAVDFLQGFENYVAVGVSNLPYLDEL"  
14493. .15410  
/gene="p9677.5"  
14493. .15410  
/gene="p9677.5"  
/note="Belongs to the MIP family of transmembrane channel  
proteins."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAB64621.1"  
/db\_xref="GI:786301"  
/translation="MSSNDSDNDKQHTLDPGVDAVITPEQPEYKHHKPKYSRDT  
LRHFILAAVEPCGTFMFLCAVYICNVANHDVALVAAPGSPGQILIMAIIGFSV  
MPSICPGVSGGALNPAMSLICLRAVSPTRCVVMMVQSIVAGMAAGGAASAMRG  
EVLFANSLGICGRTGLPIEMRGTAIILCTIVMTAVKRETNFMALPISGLIFAH  
VALITATGTGVNPARSLGAVAARFPPIHMTIYIGLISILAMSVQQLDLYTT  
YVTAERASTKEKAQKGETSSSSAAVEV"  
complement (15542. .16012)  
/gene="p9677.12"  
complement (15542. .16012)  
/gene="p9677.12"  
/note="Similar in C-terminus to Gentamicin  
3'-acetyltransferase of P. aeruginosa (Swiss Prot.  
accession number P23181)"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAB64622.1"  
/db\_xref="GI:786308"  
/translation="MSNTSBDNITVRFTVENDKGNQRLMKSYQDFEVSFPDDLDF  
NFGRLDPNLIKMAVAVSESSSEKIGIMIFNMHTWDFKDIYINDLYVDNSHVK  
GAGGKLIQFYVDADKLGTSPYVWCTDESHRAQLLYVKKVGYAPKILYKRYG"  
complement (16937. .19570)

gene  
CDS  
/gene="p9677.13"  
complement (16937. .19570)  
/gene="p9677.13"  
/note="Similar to S. cerevisiae hypothetical protein  
HHD799 (PIR accession number S45161) and S. pombe  
Query Match 17.5%; Score 217.4; DB 8; Length 37497;  
Best local Similarity 59.4%; Pred. No. 6,4e-26;  
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;  
QY 62 GTCCCAAAAGTATATTTGCATATGAAAGGTGTGATTAAGCTTATATCGACATCAT 121  
DB 3522 GGCCAAAAGCAATATTTCTGTGACTATGATGGCTGTATAGGATTAAGACATTCCTTCA 3463  
QY 122 TATTAGAGCAATTTAAGAACCCAGTAATGATGACCGTATTAATGACAGTGGAG 181  
DB 3462 TTTTGACTGACCAATTAAGGTACATCAGGGTTTA--AGAGCATTTCACTGTGAT 3406  
QY 182 ATTGTATTAAGCATTTTTCAGAAAATCATTGGAAAACATATTTGATCATTTCCG 241  
DB 3405 AGTGTCGCAAAATCTCTTCAAAAAGTCACTTAGAGACACTTGATAGCATTCTG 3346  
QY 242 AAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCCAAACACTTGA 301  
DB 3345 ATACGAAACCATTCATTTCTTATTTGTGAAAAAGAGTGAAGCTCCGACAACTGA 3286  
QY 302 AAAAGCATGAATCACCCTACAAATCATTTTAAATGTACATTTGAAAATTTGCAGAA 361  
DB 3285 AGGACACAGAAAGTAAGCAATACCAATCTTTCATTTGTCCAGAGAAAGATGCACCTCC 3226  
QY 362 CATTTATTAACATCATCTTTTACAGATCATATATTTATCTGTTCATGAAAAACATTA 421  
DB 3225 GATTCTACAGCATCCACATTTAAGGACATATTTATCTGTTCATTAACAATACTA 3166  
QY 422 CGGTAAACATGTAAATTAAGTTTCACTGCACCTTCAAAATTAGACAAACATTAAT 481  
DB 3165 CCGTCCACACATGCAATTAAGCTTCAAGAGGCATATAGGCTAAAGAAATCACATTTCT 3106  
QY 482 AACATCA-----TGGTGTATCTCTGCTTATTAATGTATCATCTCTGTTGTTTAAA 535  
DB 3105 AACATCAGATCTGTGAGTAAGAAATCTTTCACAAATGATCTTTGCTGTTGTGCAAG 3046  
QY 536 ATTTCAAACTGGTGAATTAATTAATCATTTTCAATTAACAACTGCATCCAAAATTAAT 595  
DB 3045 AGTTTGATATGTGTACATATTCGATATTAATAAATGATCATCTTAAATTAAT 2986  
QY 596 GTCCATAATGTGTAAGGTGTGTGTGGAAGAAAAGATTATCTTACATATGTTAATGC 655  
DB 2985 GTCTATTTGTGACAAACATGTGTGGGGGAAAATGTTTACAAATGCACATGATATTC 2926  
QY 656 ATGATGATTCATCATGATCAAAATATGCACTTGTGATTAATGTGATGTGGGAAATTCG 715  
DB 2925 ATACACACCTCAGTACGTAACCAAAATTTGAAAGGTGATATATGCTGATATGTCTTTT 2866  
QY 716 CAAAGAAAATGAATTAAGTTGAACATTAT 744  
DB 2865 CTAGAAACATGATCTTCTCAGCATTTAT 2857  
RESULT 12  
YSCR026A/c 2133 bp DNA linear PLN 27-Apr-1993  
LOCUS YSCR026A 2133 bp DNA linear PLN 27-Apr-1993  
DEFINITION S.cerevisiae RNA polymerase II epsilon subunit (RPO26) gene, complete  
cde.  
ACCESSION M33924  
VERSION M33924.1 GI:172452  
KEYWORDS RNA polymerase II; c-myc proto-oncogene; transmembrane protein;  
tyrosine kinase.  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
REFERENCE 1 (bases 1 to 2133)

AUTHORS Archambault, J., Schappert, K.T. and Friesen, J.D.  
TITLE A suppressor of an RNA polymerase II mutation of *Saccharomyces cerevisiae* encodes a subunit common to RNA polymerases I, II, and III  
JOURNAL Mol. Cell. Biol. 10 (12), 6123-6131 (1990)  
MEDLINE 91061718  
PUBMED 2247052  
COMMENT Original source text: *S. cerevisiae* single-copy plasmid DNA, clone 2.  
Draft entry and computer-readable sequence for [unpublished (1990)] kindly submitted by J.D. Friesen, 01-MAY-1990  
The Hospital for Sick Children  
555 University Avenue  
Toronto, Ontario, CANADA M5G 1X8  
e-mail: jfmesick@sickkids.toronto.edu.  
Location/Qualifiers  
1. 2133  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
join(1026..1045,1122..1569)  
/note="RNA polymerase II sixth subunit (RP026)"  
/codon\_start=1  
/protein\_id="AAA34989.1"  
/db\_xref="GI:172453"  
/translation="MSDYEAFAFDNGNENFDPVHFSDSEYEEKPKDFKQGETTDAN  
GRTIVGNGPDPFOHBOIRKTLKEKAIPODORATPYMTKYEARIIGRAIQIS  
NNAPVVDIEGETDPLRTMKLEAKKIPVIRRLPDPSPFDMSEVERIYDL"  
<1026..1045  
/note="RNA polymerase II sixth subunit (RP026)"  
/number=1  
1046..1121  
/note="RP026 intron A"  
1122..>1569  
/note="RNA polymerase II sixth subunit (RP026)"  
/number=2  
ORIGIN Chromosome XVI.  
Query Match 16.1%; Score 199.2; DB 8; Length 2133;  
Best Local Similarity 61.4%; Pred. No. 1.1e-22;  
Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;  
QY 171 TACAGTGAAGCATTTGTATTAAGCATTTTTCAGAAAATGCAATTTGGAAACATATTGT 230  
DB 556 TTCAAGTGATAGTGCAAAATCCCTGTTAAAAGAGTCACTTAGAGACACTGTA 497  
QY 231 ATCACATTCGGAATAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCG 290  
DB 496 TACGCAATTCGTACGAAACCAATTCATTTCTTAATGTGAAAAGAGTACGACTCG 437  
QY 291 ACAACACTGAAAAGCATGAATACCGCATCAAAAGCATTTAAATGTAATTGAAAA 350  
DB 436 CCGCACTGAAGCGACGAAGTAAGCAATCCATTTTCATTTGTCAGAGAGAG 377  
QY 351 TTGTCAAGAAGCATTTTATTAACATCTTTAAGACATCATATATTAATCTGTTATGA 410  
DB 376 ATGCAACCTCCGATTTACAGCATCCAAATTAAGGCAATATTTATCTGTTATTT 317  
QY 411 AAAAAATTAAAGTGTAAACATGTAATAAGTTTCACTCGACCTTCAAAATTAGACA 470  
DB 316 ACATAAACTTAACCTGTCACACTGCAATTAAGCTTCAGAGGCCATATAGGCTAAGAA 257  
QY 471 ACATAAATTAAACATCA-----TGGTGATCTCTGCTTATCAATGTGATCATCTCG 524  
DB 256 TGCATTTCTTAACATCAAGATCTGAGTAGAAAAATCCCTTACCAATGTAATCTTGTCTG 197  
QY 525 TTGTTTAAAAATTTCCAACTGTGCTAGTATTACATTTCAATATAAAACAATCGATCC 584  
DB 196 TTGTTGCAAGAAGTTTCGATATGTCACATTTGCAATGCAATATAAAAGATCATCC 137  
QY 585 AAAACTTAATGTCTTAAATGTGTAAAGGTTGTGTGGAAAAAAGTTTATCTTCACA 644

DB 136 TAAATTAATAGTCTTATTTGTAGCAACATGTGTGGGGAAAAATGTTTACAAATGCA 77  
QY 645 TATGTTAAGTATGATGATGATTTTACCATGATGATCAAAATATGACTTGTGATTAATGATGT 704  
DB 76 CATGATTTATTCATGACGACTCACTAGTAAACCAAAATTTGAAAGTCTATATATGTCCTGA 17  
QY 705 GGGGAAATTTGCAA 718  
DB 16 TATGCTTTTCTTA 3  
RESULT 13  
CNS061NG 797 bp DNA linear STS 30-NOV-2001  
LOCUS T7 end of clone AS09A022B07 of library AS09A from strain CLIB 533  
DEFINITION of *Saccharomyces bayanus*, sequence tagged site.  
ACCESSION AL400514  
VERSION AL400514.1 GI:12156635  
KEYWORDS STS.  
SOURCE *Saccharomyces bayanus*  
ORGANISM *Saccharomyces bayanus*  
Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;  
*Saccharomycetales*; *Saccharomycetaceae*; *Saccharomyces*.  
REFERENCE 1 (bases 1 to 797)  
Souleir, J.L., Aigle, M., Attiguenave, F., Blandin, G.,  
Bolotin-Fukuhara, M., Bon, E., Broctier, P., Casaregola, S.,  
de-Montigny, D., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S.,  
Saurin, W., Tekale, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
11152876  
JOURNAL 2 (bases 1 to 797)  
MEDLINE Bon, E., Neuvéglise, C., Casaregola, S., Attiguenave, F., Wincker, P.,  
Aigle, M. and Durrens, P.  
Genomic exploration of the hemiascomycetous yeasts: 5.  
*Saccharomyces bayanus* var. *uvaurum*  
FEBS Lett. 487 (1), 37-41 (2000)  
20584715  
TITLE 11152880  
JOURNAL 3 (bases 1 to 797)  
MEDLINE 11152880  
PUBMED 3 (bases 1 to 797)  
REFERENCE 2 (bases 1 to 797)  
Direct Submission  
Genoscope.  
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,  
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
sequ@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This STS is part of a random genomic sequencing program of thirteen  
yeast species: *Saccharomyces bayanus* var. *uvaurum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sobolofphila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.  
Location/Qualifiers  
1..797  
/organism="Saccharomyces bayanus"  
/mol\_type="genomic DNA"  
/strain="CLIB 533"  
/variety="uvaurum"  
/db\_xref="taxon:4931"  
/clone="AS09A022B07"  
/clone\_11b="AS09AA"  
/note="end : 797"  
FEATURES  
source  
misc\_feature  
/note="similar to *Saccharomyces cerevisiae* ORF YPR186c [P271 ; TP11A (transcription initiation factor) ]"  
/evidence="not\_experimental"

ORIGIN

Query Match 15.6%; Score 193; DB 11; Length 797;  
Best Local Similarity 59.3%; Pred. No. 1.4e-21;  
Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;

QY 178 GACGATTTGATTAAGCATTTTTCAGAAAATCATTGGAAACACATATTGTATCAT 237  
DB 83 GACAAATTTGCAAAATCCTTTGTTAAABAAGATCATATAGAGGACCTGTTTGGCAT 142  
QY 238 TCCGAAAAAAACCATTTCCATTGTCAGTGTGTGTAAAGGGTTAATTCGACAAAC 297  
DB 143 TCTGATACGAAGCATTTCCATGTTCTTATTTGTGGCAAGGGGGTGAACAATCCAGCA 202  
QY 298 TTGAAGAAGATGAATTCACCAATCACAAGTCATTTAAATGTACATTTGAAATTCGCA 357  
DB 203 TTGAAGGCGCAGAAAGTTACACAAATCTTTGTTGTCCGGAAGTGAATGTGAT 262  
QY 358 GAAGCATTTTAAATCAATCATCTTTTAAAGACATCATATATTCTGTCAGAAAAACA 417  
DB 263 CTCGATTTTATAGACACCTCAATTAGGGCTCATTTCTATCTGTCCACCTAGCAAA 322  
QY 418 TTAACGTAAACAATGTAAATTAAGTTTCACTGACCTTCAAAATTAAGCAACAATAA 477  
DB 323 TTAACCTGTCCAAATGTAAATTAATTTTCAAGAACCTTACAGGCTTAAGGAATCATATC 382  
QY 478 TTAACATCATATGT-----GGATCTCCCTGTTTATCATATGTATCATCTGTTGTTT 531  
DB 383 TCTAAACATCATATCTCTGAAGGTGTAAATCCGTATCAATGTATCTTCCGTGCTGTGT 442  
QY 532 AAAATTTCCAACTTGTCAGTATTTACATTTTCAATTAACAACATGATCCAAACT 591  
DB 443 ACAGATACCGTATATGTATGATGATGATGATGATGATGATGATGATGATGATGATG 502  
QY 592 AAATGCTTAATGTGTAAAGGTGTGTGTGGAAAAAGTTTATCTTACATATGTTA 651  
DB 503 AAATGCCCCATTGCAACAAGCCTTGTGTAGGGGAAATGTCTACAAATGATATGATC 562  
QY 652 AGTCATGATGATTTTACATGATCAAAATATGACATTTGTATTTGTATGATGATGATG 711  
DB 563 ATTCAATGATGAATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 622  
QY 712 TTGCAAGAAAGAAATGATATGATGATGATGATGATGATGATGATGATGATGATG 756  
DB 623 TTCTTGAGAAAGCAACACCTCTTACTACTACCAATGACGAT 667

RESULT 14  
AE016818 01  
WPCOMMENT

Sequence split into 16 fragments LOCUS AE016818 Accession AE016818

Fragment Name	Begin	End
AE016818_00	1	110000
AE016818_01	100001	210000
AE016818_02	200001	310000
AE016818_03	300001	410000
AE016818_04	400001	510000
AE016818_05	500001	610000
AE016818_06	600001	710000
AE016818_07	700001	810000
AE016818_08	800001	910000
AE016818_09	900001	1010000
AE016818_10	1000001	1110000
AE016818_11	1100001	1210000
AE016818_12	1200001	1310000
AE016818_13	1300001	1410000
AE016818_14	1400001	1510000
AE016818_15	1500001	1519138

Continuation (2 of 16) of AE016818 from base 100001 (AE016818 Eremothecium gossypii ATCC

Query Match 9.7%; Score 120.6; DB 8; Length 110000;  
Best Local Similarity 50.1%; Pred. No. 1.9e-10;  
Matches 386; Conservative 0; Mismatches 374; Indels 11; Gaps 3;

QY 25 TCGATATCATCTTTAATATCTTCTCTTATCATCAAGTCCAAAAAGTATATTGACA 84  
DB 17349 TCGTTTTCAGATTTTAAGTTCACGCGGTGAGACACCGCGGCAAAACGATTTCTGGAC 17408  
QY 85 TATGAAGGTTGTATTAAGCTTAATTCACATCATTTATAGACAAACA-TTTAAGAAC 143  
DB 17409 TACGAGGGCTGCTACAAAGGGTTTCAAGCGCGCTCCCTGTGTACGAGACACGAGACG 17468  
QY 144 CCACAGTAAATGATCGACCGTATTAATGTACATGTGACGATTTGTATTAAGCATTTTTCAG 203  
DB 17469 GCGCACCAAGGGCATTCAGGGCATTCACAGTGC---GAGCATGCGGGCGCGGTTTACAA 17524  
QY 204 AAAATCACTTTGGAAGAACATATTTGTATCACTTCGAAAAAAACCAATTCATTTGTC 263  
DB 17525 GAAGTGCCTTTGAGCGGACCTGTTCTTGACCTTGAGACAAAGCGTTTCAGCTGCAC 17584  
QY 264 AGTGTGTGTTAAAGGGTTAATTTCTGACAAACATTTGAAAGACATGAAATCCACATAC 323  
DB 17585 GGTATGTGCGAAGGGGTGACACCGCGGACGACATGAGCGGCGACAGATTCACACAC 17644  
QY 324 AAAATCATTTAATGTATCATTTGAAATTTGTCAAGACATTTTATTAATCATCATCTTT 383  
DB 17645 AAAGTGTTCAGATGCGCCGACAGAGGCTGCGGCGAAGGTTCTTCAAGACACCGCAGTT 17704  
QY 384 AAGACATCATATTTATTTCTGTTATGAAAAAAACATTAAGGTATTAATTAAGT 443  
DB 17705 GCGTTGCAAGTGTGCGGCTGCGGTGACAGCAAAAGTACAGTGTGACACATGTACAAAGC 17764  
QY 444 TTCACTGACCTTCAAAATTAAGCAACAATTAATTAACATCATGTTGATCTCC--- 500  
DB 17765 GTTCCAAAGGCTATACCTTTTGAACACACATGCGCAAGATTCAGGGCTGATGCA 17824  
QY 501 ---TGCTTATCAATGTATCATCTGTTTAAATTTTCAAACTTGTGATGAT 557  
DB 17825 GTTCCGATTCATATGACAGCAAGCGCGGTGCGTATGAGACGTTGATGAGTCTGCGCT 17884  
QY 558 ACAATTTCTATTAACAACTGATTCACAACTTAATGTTCTTAATGTGTAAAGTTG 617  
DB 17885 GCAGACACCTTCAACAGCAACCAAGCTGATGATGATGATGATGATGATGATGATG 17944  
QY 618 TGTGGAAGAAAAAGTTTATTTTCAATATGTTAAGTATGATGATGATGATGATGATG 677  
DB 17945 CGTGGGAGAGCGGACCTGATGATGATGATGATGATGATGATGATGATGATGATG 18004  
QY 678 AATATGACTTGTATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 737  
DB 18005 AAATGGAAGTGTCTGTGTGTGATGATGATGATGATGATGATGATGATGATGATG 18064  
QY 738 ACATTTAATATCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 788  
DB 18065 GAATTAATGAAGAAACCAAGGATTCATTCACCAAGAAATGATGATGATGATG 18115

RESULT 15  
AC138126/c 169063 bp DNA 1linear PRI 17-DEC-2002

LOCUS AC138126 Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.  
DEFINITION AC138126  
ACCESSION AC138126  
VERSION AC138126.1 GI:27151357  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 169063)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
TITLE Unpublished  
JOURNAL  
RECORD 2 (bases 1 to 169063)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
TITLE Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell



## COMMENT

Drive, Walnut Creek, CA 94598, USA  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.

## FEATURES

source

1..169063  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="RP11-274A19"

## ORIGIN

Query Match 8 2%; Score 101; DB 9; Length 169063;  
Best Local Similarity 50.8%; Pred. No. 2.4e-07;  
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;  
  
QY 7 GAAAGTGACGAAACCAATCGATATCATCTTTAATATCTTCTTCATCACGTCCTCC 66  
DB 80134 GAATGGGCAAAAGCTTTTAAACGAGTCCTCAATCCTTACTACACATAGAGAAATCTACT 80075  
QY 67 AAAAGTATATTGCACATATGAGGGTGTGATAAAGCTTATATGACCATCATATTATTA 126  
DB 80074 GGAGGAAATCCTACAAATGTGAAGATGTGGCAAAAGCTTCTATGATCTCAAAACTT 80015  
QY 127 GAGCAACATTTAAGAACCCAGTATGATGACCGTATAATGTACAGTGAAGCATTTGT 186  
DB 80014 ACTGAACATTAAGAAATTTATCTGTGAGAGAAACCTTACACAGT-----GAGAGATGT 79961  
QY 187 GATTAAGCATTTTTTCAGAAATTCACATTTGAAAACATATTTGATTCACATTCGAAAAA 246  
DB 79960 GGCAAAAGCTTTAACCATCTCTACATCTGTCTACACATAGATATTCATCTGAGAG 79901  
QY 247 AAACCATTCATTTGCTGAGTGTGTGTAAGGGGTTAATTCGACACACTTGAAGA 306  
DB 79900 AAACCTTACCAATGTGAAGATGTGTAAGGCTTTAACAGTCTCAACCTTACTAGA 79841  
QY 307 CATGAATACCCATACCAAGTCAATTAATGTACATTTGAAATTTGTCAAGAAGCATTT 366  
DB 79840 CATTAAGAAATTCATACCTGAGAGAAACCTTACCAATGTGAAAAAATGTGCAAAAGCTTTT 79781  
QY 367 TATTAACATCAATCTTTAA--GACATCATATATTATCTGTCATGAAAAAATTAACG 423  
DB 79780 AACCACTCTCAAACTTACTGACATTAAGAAATTCATCTGCTGAGAACTTACAAA 79721  
QY 424 TGTAAACAATGTATTAAGTTTCACTGACCTTCAAAATTAAGACACATTAATTAATA 483  
DB 79720 CCTAAAGATGTAAAGATTTTGAACACTTCAAAGTTTCTTAAACATTAAGAAAT 79661  
QY 484 CATCATGTGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAATTTCCAA 543  
DB 79660 TATGCTGAGTAATCTTGAAGATGTGAAGATGTAAACAAACCTTTAAAGTTGTAC 79601  
QY 544 ACTTGGTCAGTATTAATTAATTAATAACAACATGATCAAAACT 590  
DB 79600 ACTTGAATTTGTCAATAGATTAATTCATTAATAAAACCTTACAAGT 79554

Search completed: February 9, 2005, 05:36:23  
Job time : 5295.91 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:43:04 ; Search time 642.278 Seconds  
(without alignments)  
11419.610 Million cell updates/sec

Title: US-09-831-804-2  
Perfect score: 1239  
Sequence: 1 atgagctgaagagcagcaaac.....catcagtgattctctgataa 1239

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1239	100.0	2060	3	AAAI5398	AAAI5398 DNA encod
2	1235.8	99.7	1239	6	ABZ32205	ABZ32205 Candida a
3	99.4	8.0	2132	10	ADAA3158	ADAA3158 Human cod
4	96.8	7.8	1705	9	ACA98970	ACA98970 cDNA enco
5	96.6	7.8	2320	10	ACA56456	ACA56456 Human sig
6	96.6	7.8	2320	12	AD156252	AD156252 Human pol
7	95.2	7.7	4563	5	AA931317	AA931317 DNA encod
8	95	7.7	981	12	ADN04717	ADN04717 Antipsoi
9	94.6	7.6	2557	8	AAI51569	AAI51569 Human nuc
10	92.4	7.5	3639	5	AA864586	AA864586 DNA encod
11	92	7.4	994	12	ACH87568	ACH87568 Human gen
12	92	7.4	2873	13	ADRI4368	ADRI4368 Human NF-
13	91.8	7.4	1533	12	ADN99064	ADN99064 Novel hum
14	91.8	7.4	1533	12	ADN00633	ADN00633 Novel hum
15	91.8	7.4	2110	10	ADDB3824	ADDB3824 Human cDN
16	91.8	7.4	2760	10	ADCS6695	ADCS6695 Human mac
17	91.8	7.4	3078	10	ADBE2468	ADBE2468 Human cDN
18	91	7.3	3755	13	ADRO7315	ADRO7315 Full leng
19	91	7.3	4209	13	ADRO8223	ADRO8223 Full leng
20	90.4	7.3	2026	8	AA55863	AA55863 Human nuc

21	90.4	7.3	2114	10	ADA53124	ADA53124 Human cod
22	90.2	7.3	2230	10	ADA52931	ADA52931 Human cod
23	89.8	7.2	590	12	ACH80103	ACH80103 Human gen
24	89.2	7.2	2064	10	ADCS0762	ADCS0762 Human nov
25	89.2	7.2	2221	11	ADM01344	ADM01344 Human cDN
26	89.2	7.2	2622	5	AA668872	AA668872 DNA encod
27	89.2	7.2	2729	4	AAH16178	AAH16178 Human cDN
28	89.2	7.2	3839	6	ABK83826	ABK83826 Human cDN
29	89.2	7.2	3839	12	ADN04207	ADN04207 Antipsoi
30	89.2	7.2	3839	13	ADP25849	ADP25849 Breast ca
31	89	7.2	1991	13	ADS09824	ADS09824 Human the
32	89	7.2	2298	4	AAH16608	AAH16608 Human cDN
33	89	7.2	2299	13	ADS09825	ADS09825 Human the
34	89	7.2	2905	5	AA666143	AA666143 DNA encod
35	89	7.2	3003	12	ADN06503	ADN06503 Antipsoi
36	89	7.2	3020	5	AA92560	AA92560 DNA encod
37	89	7.2	3502	4	AA157845	AA157845 Human pol
38	88.6	7.2	1335	12	ADN98906	ADN98906 Novel hum
39	88.6	7.2	1335	12	ADN00475	ADN00475 Novel hum
40	88.6	7.2	1466	12	ACH91879	ACH91879 Human gen
41	88.6	7.2	1875	10	AD160591	AD160591 Secreted
42	88.6	7.2	2597	8	AD55855	AD55855 Human nuc
43	88.6	7.2	4227	9	ACA98938	ACA98938 cDNA enco
44	88.4	7.1	1410	12	ACH87337	ACH87337 Human gen
45	88.4	7.1	1952	10	ADC58104	ADC58104 Zinc fing

# ALIGNMENTS

RESULT 1	AAAI5398	standard; DNA; 2060 BP.
ID	AAAI5398	
AC	AAAI5398;	
XX		
XX		
DT	04-SEP-2000	(first entry)
XX		
DE	DNA encoding a transcription factor designated CATP11A.	
XX		
KW	Transcription factor; CATP11A; DNA-binding protein;	
KW	ribosomal RNA 5S gene; fungal infection; ss.	
XX		
OS	Candida albicans.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	720..1958
FT		/*tag= a
FT		/transl_except= (pos: 1296..1298, aa: Ser)
FT		/transl_except= (pos: 1734..1736, aa: Ser)
XX		
PN	WO200028037-A1.	
XX		
PD	18-MAY-2000.	
XX		
PF	09-NOV-1999;	99WO-FR002739.
XX		
PR	10-NOV-1998;	98FR-00014147.
XX		
PA	(HMRI ) HOECHST MARION ROUSSEL.	
XX		
PI	Bordon-Pallier F, Camier S, Sentenac A;	
XX		
DR	WPI; 2000-376549/32.	
XX		
DR	P-PSDB; AAY93316.	
XX		
PT	New nucleic acid encoding Candida albicans transcription factor, useful	
XX		
PT	e.g. in screening for antifungal agents and for immunization.	
XX		
PS	Claim 4; Page 32-33; 45pp; French.	
XX		
CC	The present sequence encodes a Candida albicans transcription factor, designated CATP11A. The polypeptide is a DNA-binding protein, which is	

CC Involved in initiating transcription of the ribosomal RNA 5S gene. The  
CC polynucleotide is used to screen for its specific inhibitors, potentially  
CC useful as antimycotic agents, to raise an antibody response that is  
CC protective against fungal infection and to raise antibodies. Such  
CC antibodies, as well as the polypeptides and polynucleotides are used in  
CC compositions for diagnosing and treating fungal infections, e.g. by  
CC detecting polymorphisms and mutations

XX Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 1239; DB 3; Length 2060;

Best Local Similarity 100.0%; Pred No.3.5e-234; Mismatches 0; Indels 0; Gaps 0;

Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTGAAGTGCAGCAACCAATGCATATCTTATATCTTCTTCTGATCA 60  
DB 720 ATGAGTGAAGTGCAGCAACCAATGCATATCTTATATCTTCTTCTGATCA 779  
QY 61 CGTCCCAAAAGTATATTTTGACATATGAAAGGCTGATPAAAGCTATATGACCATCA 120  
DB 780 CGTCCCAAAAGTATATTTTGACATATGAAAGGCTGATPAAAGCTATATGACCATCA 839  
QY 121 TTATTTAGACCAATTTAAGAACCCACATTAATGACGACCGTATTAATGTACAGTGAC 180  
DB 840 TTATTTAGACCAATTTAAGAACCCACATTAATGACGACCGTATTAATGTACAGTGAC 899  
QY 181 GATTTGATTAAGCATTTTTCAGAAATGACATTTTGAACACATATTTGATCATTCC 240  
DB 900 GATTTGATTAAGCATTTTTCAGAAATGACATTTTGAACACATATTTGATCATTCC 959  
QY 241 GAAAAAACAATTCATTTGTCAGTGTGTAAAGGGGTAAATCTCGACACACTTG 300  
DB 960 GAAAAAACAATTCATTTGTCAGTGTGTAAAGGGGTAAATCTCGACACACTTG 1019  
QY 301 AAAAGCATGAAGTACCCATCAAAAGTATTAATGTAAATTTGTCAGAA 360  
DB 1020 AAAAGCATGAAGTACCCATCAAAAGTATTAATGTAAATTTGTCAGAA 1079  
QY 361 GCATTTTAAACATCAATCTTTAAGACATCATATATATCTGTTGATGAAAAACATTA 420  
DB 1080 GCATTTTAAACATCAATCTTTAAGACATCATATATATCTGTTGATGAAAAACATTA 1139  
QY 421 ACGTGAACAATGTAATAAGTTTCACTCGAAGCTTCAAAATTTAGACACATTAATTA 480  
DB 1140 ACGTGAACAATGTAATAAGTTTCACTCGAAGCTTCAAAATTTAGACACATTAATTA 1199  
QY 481 AAAACATGATGTCGATCTCTGCTTATCATATGATCATCTGCTGTTTAAATTTTC 540  
DB 1200 AAAACATGATGTCGATCTCTGCTTATCATATGATCATCTGCTGTTTAAATTTTC 1259  
QY 541 CAAACTTGGTCAGTATTAACAATTTTCATATTAACAATGTCATCAAACTTAATGTCCT 600  
DB 1260 CAAACTTGGTCAGTATTAACAATTTTCATATTAACAATGTCATCAAACTTAATGTCCT 1319  
QY 601 AAATGTGTAAAGTTGTGTGGAAAAAGGTTTATCTTCATATATGTTAAAGTCATGAT 660  
DB 1320 AAATGTGTAAAGTTGTGTGGAAAAAGGTTTATCTTCATATATGTTAAAGTCATGAT 1379  
QY 661 GATTTCAACATGATCAAAATATGAGATCTTGATATATGATGTGGGAAATTTGCAAG 720  
DB 1380 GATTTCAACATGATCAAAATATGAGATCTTGATATATGATGTGGGAAATTTGCAAG 1439  
QY 721 AAAAATGAATTAAGTGAACATTAATATCTTCATGATGTAATATCCCTGATGATTA 780  
DB 1440 AAAAATGAATTAAGTGAACATTAATATCTTCATGATGTAATATCCCTGATGATTA 1499  
QY 781 TTAAAGGAACCTGAAGTGAAGAAATTTAGAACCTATTAAGATCAAGATCGAAATTAAT 840  
DB 1500 TTAAAGGAACCTGAAGTGAAGAAATTTAGAACCTATTAAGATCAAGATCGAAATTAAT 1559  
QY 841 AATTTGATGATTTAGAAACAGAGAAATTTAAAGTGAAGAGATGAAGATGAAGAA 900  
DB 1560 AATTTGATGATTTAGAAACAGAGAAATTTAAAGTGAAGAGATGAAGATGAAGAA 1619

QY 901 GATAGCTTAGATGAAAAAGAGTGTATGATCAGACTCAATGTCAAGTCAAGATCA 960  
DB 1620 GATAGCTTAGATGAAAAAGAGTGTATGATCAGACTCAATGTCAAGTCAAGATCA 1679  
QY 961 ATAAATCATTTTACTGCTTCTTTGGAAGTTCAAAGAGTGTTCCTTAACTTATTTGTAAT 1020  
DB 1680 ATAAATCATTTTACTGCTTCTTTGGAAGTTCAAAGAGTGTTCCTTAACTTATTTGTAAT 1739  
QY 1021 AGTGGAGAGAGATCAATTTGCTCTAAGATATATTTGATAGATGATTTTCTAGAGATAT 1080  
DB 1740 AGTGGAGAGAGATCAATTTGCTCTAAGATATATTTGATAGATGATTTTCTAGAGATAT 1799  
QY 1081 GATTTACGTGACATTTGAAATGTCATGATGATATTTTACAAAGATTAAGTCAATTCCTTA 1140  
DB 1800 GATTTACGTGACATTTGAAATGTCATGATGATATTTTACAAAGATTAAGTCAATTCCTTA 1859  
QY 1141 AATAGTATAGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCGAGATG 1200  
DB 1860 AATAGTATAGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCGAGATG 1919  
QY 1201 GATTTATGCGCAAAATGAAGACATCAGTGAATTTCTGATTA 1239  
DB 1920 GATTTATGCGCAAAATGAAGACATCAGTGAATTTCTGATTA 1958

RESULT 2  
AB332206  
ID AB332206 standard; DNA; 1239 BP.  
XX AB332206;  
XX AC AB332206;  
XX DT 30-JAN-2003 (first entry)  
XX DE Candida albicans essential gene SEQ ID NO 6493.  
XX KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
XX KW signal transduction; DNA replication; cell division; growth;  
XX KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.  
XX OS Candida albicans.  
XX PN WO200253728-A2.  
XX PD 11-JUL-2002.  
XX PP 26-DEC-2001; 2001WO-US049486.  
XX PR 29-DEC-2000; 2000US-0259128P.  
XX PR 20-FEB-2001; 2001US-0079202P.  
XX PR 22-AUG-2001; 2001US-0314050P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Roemer T, Jiang B, Boone C, Buesey H, Ohlsen KL;  
XX DR WPI: 2002-566694/60.  
XX DR P-PsDB; ABR73656.  
XX PT Constructing strains for identifying gene products as effective targets  
XX PT for therapeutic intervention, by inactivating in the strain one allele of  
XX PT a gene and placing other allele of the gene under conditional expression.  
XX PS Claim 37; SEQ ID NO 6493; 167bp + Sequence Listing; English.  
XX CC The invention relates to constructing (M1) a strain of diploid fungal  
XX CC cells in which both alleles of a gene are modified, comprising modifying  
XX CC one allele by insertion or replacement by a cassette having an  
XX CC expressible selectable marker and modifying other allele by  
XX CC recombination, of a promoter replacement fragment with a heterologous  
XX CC promoter, so that expression of the second allele is regulated by the  
XX CC promoter. (M1) is useful for constructing a strain of diploid fungal  
XX CC cells in which both alleles of a gene are modified. The diploid fungal

cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of *C. albicans* cells and for treating infection by *C. albicans*. The present sequence is that of an essential *Candida albicans* gene used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office

Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 99.7%; Score 1235.8; DB 6; Length 1239;

Best Local Similarity 99.8%; Pred. No. 1.4e-233;

Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 ATGAGTGAAGTACGAAACCAATCGATATCATCTTAAATCTTCTTCTTCATCA 60
DB 1 ATGAGTGAAGTACGAAACCAATCGATATCATCTTAAATCTTCTTCTTCATCA 60
QY 61 CGTCCAAAAAGTATATTTGACATATGAAAGGTTGATTAAGCTTAATGCAACATCA 120
DB 61 CGTCCAAAAAGTATATTTGACATATGAAAGGTTGATTAAGCTTAATGCAACATCA 120
QY 121 TTATTTAGAGCAACTTTAAAGACCCAGATATGATGACCGTATTAATGTAACAGTGCAC 180
DB 121 TTATTTAGAGCAACTTTAAAGACCCAGATATGATGACCGTATTAATGTAACAGTGCAC 180
QY 121 TTATTTAGAGCAACTTTAAAGACCCAGATATGATGACCGTATTAATGTAACAGTGCAC 180
DB 121 TTATTTAGAGCAACTTTAAAGACCCAGATATGATGACCGTATTAATGTAACAGTGCAC 180
QY 181 GATTGTGATTAAGCAATTTTCAGAAATTCACATTTGGAACACATATTTGATCATCTTC 240
DB 181 GATTGTGATTAAGCAATTTTCAGAAATTCACATTTGGAACACATATTTGATCATCTTC 240
QY 241 GAAAAAAAACCATTTCCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACAACCTTG 300
DB 241 GAAAAAAAACCATTTCCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACAACCTTG 300
QY 301 AAAAGACATGAATTCACCCATACAAAGTCATTTAATGTACATTTGGAATTTGCAAGAA 360
DB 301 AAAAGACATGAATTCACCCATACAAAGTCATTTAATGTACATTTGGAATTTGCAAGAA 360
QY 361 GCAATTTTAAACATCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTA 420
DB 361 GCAATTTTAAACATCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTA 420
QY 421 ACGTGTAAACATGATTAAGTTTCACTCGACCTTCAAAATTTAGCAACAATTAATTA 480
DB 421 ACGTGTAAACATGATTAAGTTTCACTCGACCTTCAAAATTTAGCAACAATTAATTA 480
QY 481 AAACATCATGATGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAATTTTC 540
DB 481 AAACATCATGATGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAATTTTC 540
QY 541 CAAACTGTGTCAATTAACAATTCATATAAACAATGATCCAAAACCTTAATATGTCCT 600
DB 541 CAAACTGTGTCAATTAACAATTCATATAAACAATGATCCAAAACCTTAATATGTCCT 600
QY 601 AAAATGTGTAAAGGTTGTGGGAAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660
DB 601 AAAATGTGTAAAGGTTGTGGGAAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660
QY 661 GATTCTTACATGATCAAAAATATGACCTTGATTTGTGATGTGGGAAAAATTTGCAAG 720
DB 661 GATTCTTACATGATCAAAAATATGACCTTGATTTGTGATGTGGGAAAAATTTGCAAG 720
QY 721 AAAAATGAATTAAGTGAACATTAATAATATCTTCATGATGTAATATCCCTGATGATTA 780

```

```

DB 721 AAAAATGAATTAAGTGAACATTAATAATATCTTCATGATGTAATATCCCTGATGATTA 780
QY 781 TTTAAAGAAACCTGAAGTGAAGAAAAATATGAGAACCTTATTGATCAAGATGGAATTAAT 840
DB 781 TTTAAAGAAACCTGAAGTGAAGAAAAATATGAGAACCTTATTGATCAAGATGGAATTAAT 840
QY 841 AATTTCATGAATTTAAGAACAGAAATTTAAAGGTGAAGAGATGAGAAATGAGAA 900
DB 841 AATTTCATGAATTTAAGAACAGAAATTTAAGGTGAAGAGATGAGAAATGAGAA 900
QY 901 GATAGCTGATGATGAAAAAGAGTGAATGTTAGTACAGACTCAATGTCAGCTCAAGATCA 960
DB 901 GATAGCTGATGATGAAAAAGAGTGAATGTTAGTACAGACTCAATGTCAGCTCAAGATCA 960
QY 961 ATTAATCATTTTACCTGCTCTTTGGAAGTTTCAAGAGTTCCTAACTTATTCGAAT 1020
DB 961 ATTAATCATTTTACCTGCTCTTTGGAAGTTTCAAGAGTTCCTAACTTATTCGAAT 1020
QY 1021 AGTGGGAGAGATGCAATTTGCTTAAGATTAATTTGATGATGATGTTTCTAGAAATAT 1080
DB 1021 AGTGGGAGAGATGCAATTTGCTTAAGATTAATTTGATGATGATGTTTCTAGAAATAT 1080
QY 1081 GATTTCAGTGCACATTTGGAATGTCATGATGATTAATTTTCAAGAAATGAGTCAATCTTA 1140
DB 1081 GATTTCAGTGCACATTTGGAATGTCATGATGATTAATTTTCAAGAAATGAGTCAATCTTA 1140
QY 1141 AATGATTAAGAAAAAGAGAACTCCAGAAAGTGAACCATTTGTTAAAAAGCCAGATG 1200
DB 1141 AATGATTAAGAAAAAGAGAACTCCAGAAAGTGAACCATTTGTTAAAAAGCCAGATG 1200
QY 1201 GATTATTTGCCAAATGAATGATCATGATTTCTCGATTA 1239
DB 1201 GATTATTTGCCAAATGAATGATCATGATTTCTCGATTA 1239

RESULT 3
ID ADA53158 standard; cDNA; 2132 BP.
XX ADA53158;
AC 20-NOV-2003 (first entry)
DT
XX
DE Human coding sequence. SEQ ID 726.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
XX Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX Inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX EP1293569-A2.
PN
XX
XX 19-MAR-2003.
PD
XX
XX 21-MAR-2002; 2002EP-0006586.
PF
XX
XX 14-SEP-2001; 2001JP-00328381.
PR
XX
XX 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX WPI, 2003-395539/38.
XX
XX P-P-SDB; ADA54797.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
XX and/or membrane proteins, useful for developing medicines for diseases in

```

PT which the gene is involved, or as target molecules for gene therapy.  
XX  
XX Claim 1; SEQ ID NO 726; 205pp; English.  
XX  
XX The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA5433-  
CC ADA54077). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;  
  
Query Match 8.0%; Score 99.4; DB 10; Length 2132;  
Best Local Similarity 50.6%; Pred. No. 4.7e-10;  
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;  
  
QY 7 GAAAGGACGAAACCAATCGATATCATCTTTAATCTCTCTCTTCATCAGCTCC 66  
DB 1384 GAATGGCGAAACCTTTAACCGATCTCTCAATCTTACTACATAGAGATTCATCT 1443  
QY 67 AAAAGTATATTTGACATATGAAAGGATGATTAAGCCTATATCGACCATCATTTA 126  
DB 1444 GGAGAGAAATCCACAAATGTGAAGATGTGCAAGCTTCTATGATCTCAAAACTT 1503  
QY 127 GAGCAACATTTAAGAACCAAGTAATGATCGACCGTATATATGACAGTGT 186  
DB 1504 ACTGAACATAGAAATTTCACTGAGAGAAACCTTACACATGT-----GAGGAATGT 1557  
QY 187 GATTAAGCATTTTGTGAAATTCATTTGAAACATATTTGATACATTCGGAATA 246  
DB 1558 GGCAGAACCTTTAACCATCTCTACACCTGTCTACATTAAGGATTAATCTACTGAGAG 1617  
QY 247 AAACCATTCATTTGTCAGTGTGTGTTAAAGGGGTAATTTCTGACACATTTGAAAGA 306  
DB 1618 AAACCTTACCAATGTGAAGATGTGTTAAAGCCTTTAACGATCTCAACCTTACTATA 1677  
QY 307 CATGAATACCCATACAAAGTCATTTAAATGTCAATTTGAAATTTGCAAGAGCATTT 366  
DB 1678 CATTAAGATTCATCATCTGAGAGAAACCTTACCAATGTGAAATTTGGCAAGCTTTT 1737  
QY 367 TATTAACATCATCTTTAA---GACATCATATATTTATCTGTCATGAAATTAACATTAAG 423  
DB 1738 AACCATCTCTCAAACTTACTGACATTAAGAAATTCATATCTGTAGAAACTTACAAA 1797  
QY 424 TGTAAACAATGTAAATTAAGTTTCACTCGACCTTCAAAATTTAGCAACATTAATTA 483  
DB 1798 CCGAAAGATGTAAAGTATTTTGAAGACCTTCAAACTTTCTAAACATTAAGAAAT 1857  
QY 484 CATCATGTGTGATCTCTGCTTATCAATGTATCATCTGTGTTTAAATTTCCAA 543  
DB 1858 TATGCTGTGAGAAATCTTGAAGATGTGAAGATGTAAACAACTTTAAAGTTGTAC 1917  
QY 544 ACTGTGTGATTTACATTTTCAATTAACAACTGATCCAAACT 590  
DB 1918 ACTGATGTGTGATTAAGATTAATTCATTAACAACTTAAGT 1964  
  
RESULT 4  
ACA98970  
ID ACA98970 standard; cDNA; 1705 BP.  
XX  
XX ACA98970;  
XX  
DT 25-JUL-2003 (first entry)  
XX  
DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.  
XX  
XX Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;  
KW anti-conventional; nootropic; neuroprotective; cerebroprotective; anti-HIV;  
KW anti-allergic; anti-inflammatory; thyromimetic; gene therapy;  
KW cell proliferative disorder; cancer; atherosclerosis;  
KW neurological disorder; epilepsy; Huntington's disease; stroke;  
KW immune disorder; inflammatory disorder; AIDS; allergy;

KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2003023003-A2.  
XX  
XX 20-MAR-2003.  
XX  
XX  
XX 05-SEP-2002; 2002MO-US028540.  
XX  
XX  
XX 07-SEP-2001; 2001US-0317792P.  
XX  
XX 07-SEP-2001; 2001US-0317912P.  
XX  
XX 14-SEP-2001; 2001US-0322270P.  
XX  
XX 21-SEP-2001; 2001US-0324040P.  
XX  
XX 28-SEP-2001; 2001US-0326732P.  
XX  
XX 19-OCT-2001; 2001US-0346715P.  
XX  
XX 25-JAN-2002; 2002US-0351749P.  
XX  
XX 22-FEB-2002; 2002US-0359498P.  
XX  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe TJ;  
PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BW;  
PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;  
PI Hafalia AJP, Sanjanwala B, Margulis JP, Gorrard AE, Lee SY, Ison CH;  
PI Baughn MR, Chawla NK, Nguyen DB, Swarnaker A, Zebardjian Y, Shah P;  
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kabile AE;  
PI Burford N, Ramkumar J;  
XX  
XX WPI; 2003-313243/30.  
XX  
XX P-PSDB; AB096722.  
XX  
XX  
XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
PT creating and preventing diseases or conditions associated with the  
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
PT infections.  
XX  
XX  
XX Claim 5, Page 340-341; 345pp; English.  
XX  
XX  
XX The invention describes a novel human isolated nucleic acid-associated  
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in  
CC diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression or overexpression of NAAP, such as cell  
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
CC disorders, or infections. These are also useful in assessing the effects  
CC of exogenous compounds on the expression of nucleic acid and amino acid  
CC sequences of NAAP. The NAAP or its fragments are useful in screening  
CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
CC or in altering the expression of the target polynucleotide and compounds  
CC that specifically bind to or modulate the activity of the polypeptide.  
CC The microarray is useful in monitoring or measuring protein-protein  
CC interactions, drug-target interactions, and gene expression profiles.  
CC This sequence encodes a novel human nucleic acid-associated protein  
CC (NAAP)  
XX  
XX  
SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;  
  
Query Match 7.8%; Score 96.8; DB 9; Length 1705;  
Best Local Similarity 52.9%; Pred. No. 1.5e-09;  
Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;  
  
QY 7 GAAAGTACGAAACCAATCGATATCATCTTTAATCTCTCTCTTCATCAGCTCC 66  
DB 695 GAATGTCCCAAAACATTTTGGATCTTTCACACCTTAACATTAAGAAATCCAAACT 754  
QY 67 AAAAGTATATTTGACATATGAAAGGATGATTAAGCCTTAATGACATCATTTA 126  
DB 755 AGATGAATTTCTACAAATGTGAAGCATATGAAAGGCTTTAACTGTCTCAACCTT 814



KW effector-like polypeptide; cancer; immunopathology; neuropathology;  
KW drug development; toxicology; carcinogenicity;  
KW signalling pathway polypeptide; adrenal gland; bladder; bone;  
KW bone marrow; breast; cervix; tumour; immunopathology; AIDS;  
KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;  
KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.

OS Homo sapiens.

PN US2004010136-A1.

PD 15-JAN-2004.

PF 26-NOV-2002; 2002US-00305720.

PR 30-JAN-1998; 98US-00016434.

XX (INCY-) INCYTE GENOMICS INC.

PI Au-Young J, Seilhamer JI;

DR WPI; 2004-090520/09.

PT New composition comprising polynucleotide probes, useful as array  
PT elements in a microarray for monitoring the expression of target  
PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic  
PT fragments.

PS Claim 6; SEQ ID NO 1054; 73pp; English.

XX The invention relates to a composition of polynucleotide probes  
XX comprising first polynucleotide probes comprising at least a portion of a  
XX gene encoding a receptor-like polypeptide, second polynucleotide probes  
XX comprising at least a portion of a gene encoding a transducing  
XX polypeptide and third polynucleotide probes comprising at least a portion  
XX of a gene encoding an effector-like polypeptide. The probes of the  
XX composition are useful as array elements in a microarray for monitoring  
XX the expression of target polynucleotides. The microarray is useful in the  
XX diagnosis and treatment of cancer, an immunopathology or a  
XX neuropathology. It can also be used for drug discovery and development.  
XX toxicological and carcinogenicity studies, forensics or pharmacogenomics.  
XX Microarrays can also be used for monitoring the progression of diseases  
XX that may be associated with the altered expression of signalling pathway  
XX polypeptides. The composition can also be used to purify a subpopulation  
XX of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile  
XX is also useful for the diagnosis and treatment of cancer, e.g. cancers of  
XX the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,  
XX an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or  
XX ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,  
XX epilepsy, Alzheimer's disease or depression. This sequence represents a  
XX human polynucleotide probe of the invention. Note: The sequence data for  
XX this patent did not form part of the printed specification but was  
XX obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html.

XX Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

XX Query Match 7.8%; Score 96.6; DB 12; Length 2320;

XX Best Local Similarity 51.3%; Pred. No. 1.7e-09;

XX Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAAGTATATTTGACATATGAAAGGTGATTAAGGCTTAATTCAGCATTTATTAGA 128

DB 964 AGAGAAACCTCAATGTAAAGATGTGTAAGCTTTTAAACCGATCTTCAACCTTAC 1023

QY 129 GCAACATTTAAGAACCCAGATATGATGACCGTATTAATGTACAGTGAGCATTTGTA 188

DB 1024 TACCCCTAGAAAATTCATATGAGAGAAACCTTCAATATG-----GAAGAATGTG 1077

QY 189 TAAAGCATTTTCAAGAAATTCATTTGAAACACATATTTGATCATTTCCGAAAAA 248

DB 1078 CAAGGCTTTAAGACAGCTCCTCAACCTTACTACATTAAGATTAATTCATTAAGAGAA 1137

QY 249 ACCATTCATGTTTCAGTGTGTGTAAGGGGTAAATTCCTGCAGAACCTTGAAAAAGACA 308

DB 1138 ACCCTTACAAATGTATAAAATGTGAAAAGCCTTTAACAGTCTGCACACCTTACACACA 1197

QY 309 TGAATTCACCCATAGAAAGTCAATTTAAATGTACATTTGAAAAATTCAGAAAGCATTTTA 368

DB 1198 TGAGGTAAATTCATCTAGAGAGAAACCTTACAAATGTAAAAATGTGAAAAGCCTTTAA 1257

QY 369 TAAACA--TCATCTTTAAGACATCATATATATATCTGTTCATGAAAAAATTAACGTG 425

DB 1258 TCATTTTCACACCTTACTACACATAGATATATTCATCTGAGAGAAAACCTTACAAATG 1317

QY 426 TAAACATGTATAAAGTTTTCATCTGACCTTCAAAATTAGACAAACATTAATTAACA 485

DB 1318 TAAAGAAATGTGTTAAGCTTTTAAACCTTCAACCTTAAACATTAAGATTAATCA 1377

QY 486 TCATGTTGATCTCTGCTTATCATATGATCATCTGTTGTTTAAATTTCCAAAC 545

DB 1378 TACTGAGAGAAAGCTTACAAATCTTAAAGAAATGTGAAAAAGCTTTTAAACCAATCCTCAAA 1437

QY 546 TTGATCAGTATTTACATTTTCATATATAAACAATGCATCCAAAACCTTAATGCTTAATG 605

DB 1438 ACTTACTGAACATTAAGAAAATTCATACCTGAGAGAAACCTTA--TGAATGTAAAAATG 1494

QY 606 TCGTAAAGGTGTGTGGGAAAAAGTTTATCTTCAATATGTTAAGTCATGATGA 662

DB 1495 TGGCAAGCTTTTAAACCAAGTCTCTCAAAATCTTATGACATTAAGAAAAGTCATACAGA 1551

RESULT 7

AAS91317  
ID AAS91317 standard; cDNA; 4563 BP.

AC AAS91317;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #27121.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HISE-) HISEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG27130.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX Claim 1; SEQ ID NO 27121; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX sequences. (I) is useful as hybridisation probes, polymerase chain

XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX and in recombinant production of (II). The polynucleotides are also used

XX in diagnostics as expressed sequence tags for identifying expressed

XX genes. (I) is useful in gene therapy techniques to restore normal



CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

80 Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;  
 Query Match 7.7%; Score 95.2; DB 5; Length 4563;  
 Best Local Similarity 52.7%; Pred. No. 3.5e-09;  
 Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

QY 7 GAAGTACGCAAAACCAATGATATCATCTTAAATCTTCTTCTTCATCAGCTCC 66  
 DB 1315 GAATGTGGCAAAAGCTTTAACAGCTCCTCACTTCAATAAACAAATTCATCT 1374  
 QY 67 AAAAGTATTTTGGCATATGAAAGGTGTGATTAAGCCATTAATGACATCATTTA 126  
 DB 1375 GAGGGGAAACCTCAAAATGTGAAGATGTGGCAAAAGCTTTTAAACATCTTTAGCCCTT 1434  
 QY 127 GAGCAACATTTAAGAACCCAGATATGATGACCGTAAATGTACAGAGCATTTG 186  
 DB 1435 ACTAAACATTAAGTATTAATCTATCTGGGAGAAACCAATCAAAATGT-----GAAATGT 1488  
 QY 187 GATTAAGCATTTTTCAGAAAATCATTTTGAACACATATTTGTATCATCTTCGAAA 246  
 DB 1489 GGGAAAGCTTTTAGGCAATCTCAACCTTACTAGACATTAAGCAATTCATCTGGAG 1548  
 QY 247 AAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATCTCGACAACTTGAAGA 306  
 DB 1549 AAACCCCTCAAAATGTGAAGATGTGGCAAAAGCTTTTAAACATTTCTCAGACCTTAGAGA 1608  
 QY 307 CATGAATCACCCATCAAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGACATTT 366  
 DB 1609 CATTAAGTATTTCTACTCTGAAAAGAAACCTTCAAAATGTGAAGATGTGGAAAAGCTTTT 1668  
 QY 367 TATTAACATCATCTTTAAGA---CATCATATATTTATCTGTGATGAAAAAAGCATTAACG 423  
 DB 1669 AGCGAAGCTTCAACCTTGAACCATCATATTAATCATCTGGAAGAAACCTTCAAA 1728  
 QY 424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTTAGCAACAATTAATTA 483  
 DB 1729 TGTGAAGAATGTGTAAAGCTTTTAAAGTGTCAATCAAACTTGTATTAAGTAAAT 1788  
 QY 484 CATCATGCTG 493  
 DB 1789 CATCTGGAG 1798

RESULT 8  
 ID ADN04717 standard; cDNA; 981 BP.  
 AC ADN04717;  
 DT 01-JUL-2004 (first entry)  
 XX Anti-psoriatic cDNA sequence #572.  
 DE  
 XX  
 KW de; gene; anti-psoriatic; gene therapy; psoriasis; diagnosis.  
 XX  
 OS Homo sapiens.  
 XX

PN WO2004028479-A2.  
 XX  
 PD 08-APR-2004.  
 XX  
 PF 25-SEP-2003; 2003WO-US030907.  
 XX  
 PR 25-SEP-2002; 2002US-0414006P.  
 XX  
 PA (GENTH ) GENTECH INC.  
 XX  
 PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PW, Wood WT,  
 PI Wu TD;  
 XX  
 DR WPI; 2004-305105/28.  
 DR P-PDSB; ADN04718.  
 XX  
 PT New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.  
 PS  
 XX Claim 1, SEQ ID NO 1111; 3069bp; English.  
 CC The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polynucleotides of the invention.

SQ Sequence 981 BP; 380 A; 178 C; 170 G; 253 T; 0 U; 0 Other;  
 Query Match 7.7%; Score 95; DB 12; Length 981;  
 Best Local Similarity 50.4%; Pred. No. 3e-09;  
 Matches 289; Conservative 0; Mismatches 275; Indels 9; Gaps 2;

QY 89 AAGGCTGTATTAAGGCTTAATCGACATCATTTATTAAGCAACATTAAGAACCCACA 148  
 DB 1 AAGAAATGTGGCAAAAGCTTTAGATATTCCTCAACCTTCAATTAAGATTAATTCATA 60  
 QY 149 GTATATATGACCGTATTAATGTATGACGTGAGCATTTGTATAAGCATTTTCAAGAAAT 208  
 DB 61 CTGGAGAGAAACCTTCAAAATGCA-----ATGAATGTGTAAAGCTTTAACTGTGCT 114  
 QY 209 CACATTTGAAACACATATTTGTATCATCTTCGAAAAAAAACCATTCATTTGTCAGTGT 268  
 DB 115 CAACTTTACTTAACATTAAGATTCATTAATCTGGAAGAAACCTTCAAAATGTGAAGAT 174  
 QY 269 GTGGTAAAGGGGTTAATTTCTGCAACACTTGAAGAAAGCATGAATCACCCATCAAAAGT 328  
 DB 175 GTGGCAAGCTTTTAAACAGTCTCAACCTTACTAGACATTAAGATTAATGATACGAG 234  
 QY 329 CATTAAATGTATCATTTGAATAATTTGCAAGAGCATTTTATTAACATCAATCTTTAAGC 388  
 DB 235 AGAAACCTTCAAAATGTGAAGATGTGAAGGCTTTTAAACGATCCACAACTCTAAAC 294  
 QY 389 ATCATATTAATTTGTTCATGAAGAAACATTAACGTTAAGCAATTAATTAAGTTTCA 448  
 DB 295 ATTAAGAAATTTATTAAGTAAAGAAACCATTAAGTAAAGATGTGAAGAAAGCCCTTA 354  
 QY 449 CTCGACCTTCAAAATTAAGCAACATTAATTAAGATCATGATGATCTCTGCTTATC 508  
 DB 355 GTGTATTTCAACCTTACTTAACATTAAGTATTAATTTACTGAGCAAGAAACCTTACAAT 414  
 QY 509 AATGTATCATCTGTGTTGTTTAAATAATTTCAAACTTGTCAGTATTAACAATTTCA 568  
 DB 415 GTGACGAATGTGCAAGTGCCTTTAGGGCATTTCAACCTTACTGAACATTAAGAGATTC 474  
 QY 569 TAAACAACGTGATCCAAACCTTAATATGCTTAATGTGTAAAGTTGTTGGAAAA 628  
 DB 475 ATACTGGAAGAAACCTTAC---AAATGAATGAATGTGTAAAGCTTTAACTGTGCT 531  
 QY 629 AAGTTTATCTTACATATGTTTAAGTCATGATG 661  
 DB 532 CAACCTTACTTAACATTAAGATTCATACG 564

RESULT 9  
AAL51569  
ID AAL51569 standard; DNA; 2597 BP.  
XX  
AC AAL51569;  
XX  
DT 10-APR-2003 (first entry)  
XX  
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.  
XX  
KW Human; gene; de; nucleic acid-associated protein; NAAP; arteriosclerosis;  
cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;  
cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;  
menetal retardation; neurological disorder; Alzheimer's disease; epilepsy;  
Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;  
Crohn's disease; transgenic animal; animal model.  
XX  
OS Homo sapiens.  
XX  
PN WO200300864-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 20-JUN-2002; 2002WO-US021179.  
XX  
PR 22-JUN-2001; 2001US-0300518P.  
PR 29-JUN-2001; 2001US-0301787P.  
PR 29-JUN-2001; 2001US-0301792P.  
PR 29-JUN-2001; 2001US-0301892P.  
PR 29-JUN-2001; 2001US-0301893P.  
PR 06-JUL-2001; 2001US-0303405P.  
PR 06-JUL-2001; 2001US-0303442P.  
PR 15-MAR-2002; 2002US-0364438P.  
XX  
PI (INCYTE) INCYTE GENOMICS INC.  
XX  
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;  
PI Arvizu CS, Ison CH, Honchell CD, Lee RA, Yue H, Forsythe IU;  
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala NM, Baughn MR;  
PI Borowsky ML, Yao MG, Wallia NK, Bandman O, Lai PG, Becha SD, Lee SY;  
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjian Y, Lu Y;  
XX  
DR MPI; 2003-201420/19.  
DR P-PSDB; AAO16419.  
XX  
PT New nucleic acid-associated proteins and polynucleotides, useful for  
diagnosing, treating or preventing cell proliferative (e.g. cancer),  
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune  
PT disorders (e.g. AIDS).  
XX  
PS Claim 12; Page 292-293; 312pp; English.  
XX  
XX The invention comprises the amino acid and coding sequences of human  
nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
the invention are useful for diagnosing, treating or preventing disorders  
associated with aberrant expression of NAAP, such as: cell proliferative  
disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis  
or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia  
or mental retardation); neurological disorders (e.g. Alzheimer's disease,  
Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders  
(e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of  
the invention are useful for creating transgenic animals to model human  
disease. The present DNA sequence encodes a human nucleic acid-associated  
protein of the invention  
XX  
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

Query Match 7.6%; Score 94.6; DB 8; Length 2597;  
Best Local Similarity 55.1%; Pred. No. 4.2e-09;  
Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;

QY 78 TTGCACATATGAGAGGGTGTGATTAAGCCTATATATGACCACCATATTAAGACAAATTT 137  
DB 1070 TTATATAATGTGAAGATGTGGCAAAAGCCTTATATGTCTCAACCTTACTCAACATTA 1129  
QY 138 AAGAACCCAGTATGATTCGACCGGTATTAATGTACATGAGCGATTGGATTAAGCAATT 197  
DB 1130 GAGAAATTCATATGAGAGAAACCTTACAAATGT-----GAAAGTGTGGCAAAAGCCTT 1183  
QY 198 TTTCAAAAAATCACAATTGGAAACACATATTTGTATCATTCCGAAAAAAACCATTTCCA 257  
DB 1164 TAACGTGTCCCTAATCTTACTCAACATTAAGAGATTCTACTGTGAGAAAAACCTATGAA 1243  
QY 258 TTGTTCACTGTGTGTGTAAGGGGTTAATTCTTCGACAACATTTGAAAGACATGAATGAC 317  
DB 1244 ATGTGAAGAATGTGGCAAAAGCCTTAAACATCTCCACACCTCACACACATTAAGAAT 1303  
QY 318 CCATCAAAAGTCATTAATGTACATTGGAATTTGTCAAGAGATTTTAA--ACA 374  
DB 1304 TCATACCGAGAGAAACCTTACAAATGTGAAGATGTGGCAAAAGCCTTAAACCAATTC 1363  
QY 375 TCAATCTTTAAGACATCATATATTTATCTGTTCATGAAAAACATTAACGTGTAACAATG 434  
DB 1364 ACACTTACTACACATTAAGATTAATTCATCTGAGAGAAACCTTCAAAATGTAAAGATG 1423  
QY 435 TAATAAGTTTCTACTGCACCTTCMAAATTAGACAACATTAATTAACATCATGTG 493  
DB 1424 TGGCAAAAGCTTTTAAGCGGTCTCAAACTTACTGAACATATGATAATTCACTACGTGAG 1482

RESULT 10  
AAS64586  
ID AAS64586 standard; cDNA; 3639 BP.  
XX  
AC AAS64586;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #390.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HUSE-) HUSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR MPI; 2001-639362/73.  
XX  
DR P-PSDB; ABB00399.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID NO 390; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptides and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64157-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;  
 Query Match 7.5%; Score 92.4; DB 5; Length 3639;  
 Best Local Similarity 54.1%; Pred. No. 1.2e-08;  
 Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;  
 QY 69 AAGATATATTTGCACATATGAGGTTGTGATTAAGCTTATATGACCATCATTTATTAGA 128  
 Db AGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTAGCCATTCTTCAACCTTGC 797  
 QY 129 GCACATTTAAGAACCCACAGTATGATTCACCGTATTAATGTAAGTGAAGCATTTGTGA 188  
 Db 798 TAAACATTAAGAGATTTATCTGGAGAGAAACCTTAAGATGT-----GAGAGATGTGG 851  
 QY 189 TAAAGATTTTTCAGAAATTCACATTTGGAACACATATTTGATCATTCCGAAAAAAA 248  
 Db 852 CAAAGCTTTTACCGCTTCTTCAACCTTGTCTAAACATTAAGATTTATCTGAGAGAA 911  
 QY 249 ACCATTCATTTGTTGAGTGTGTGTAAGAGGTTTATTTCTGACACACATTTGAAGACA 308  
 Db 912 ACCCTACAAATGTAAAGATGTGGCAAGCTTTAGCAATTCCTCAACCTTGTGATCA 971  
 QY 309 TGAATATCCCATTAAGATGATTTAATGTAATTTGAAATTTGTCAAGAGCATTTTA 368  
 Db 972 TAAAGATACCTATCTGAAGAGAAACCTTCAATGTAAAGATGTCAAAAGCTTTTAA 1031  
 QY 369 TAAACATCATTC--TTTAAGACATCATATATTTATCTGTCATGAAGAAACATTAAGCTG 425  
 Db 1032 GCGACTCTCAACCTTACTTAAACATTAATATATCATGCTGAGAGAAACCTTCAATG 1091  
 QY 426 TAAACATGTATTAAGATTTTCACTCGACCTTCAAAATTAGCAACATTAATTTAAACA 485  
 Db 1092 TGAAGATGTGGCAAGCTTTTATGATGATCTTCAAACTTACTTATCATTAAGTTTATCA 1151  
 QY 486 TCATGTGTGATCTCT 501  
 Db 1152 TACTGGAGAGAAACCT 1167  
 RESULT 11  
 ACH87568/C  
 ID ACH87568 standard; DNA; 994 BP.  
 XX  
 AC ACH87568;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Human genome derived single exon probe #20763.  
 XX  
 KW Human; probe; 86; gene expression; single exon probe; microarray;  
 XX  
 KW alternative splicing event; genomic alteration.  
 OS Homo sapiens.  
 XX  
 XX US2003194704-A1.  
 XX  
 PD 16-OCT-2003.

XX  
 PF 03-APR-2002; 2002US-00029386.  
 XX  
 PR 03-APR-2002; 2002US-00029386.  
 XX  
 PA (PENN/) PENN S G.  
 PA (RANK/) RANK D R.  
 PA (HANZ/) HANZEL D K.  
 XX  
 PI Penn SG, Rank DR, Hanzel DK;  
 XX  
 DR WPI; 2004-119264/12.  
 XX  
 PT New human genome-derived single exon nucleic acid probes useful for human  
 PT gene expression analysis, for identifying or characterizing alternative  
 PT splicing events, for assessing genomic alterations or as tools for  
 PT surveying tissues.  
 XX  
 PS Claim 1, SEQ ID NO 20763; 80pp; English.  
 XX  
 CC The invention relates to a nucleic acid probe for measuring human gene  
 CC expression, comprising any of the 27,400 fully defined nucleotide  
 CC sequences in the specification, or their complements or fragments, and  
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
 CC fully defined in the specification. The probe is a single exon probe that  
 CC hybridizes under high stringency conditions to a nucleic acid molecule  
 CC expressed in human cells or tissues. Also included are a spatially-  
 CC addressable set of single exon nucleic acid probes for measuring human  
 CC gene expression (comprising a plurality of single exon probes separately  
 CC probes cited above, where each of the plurality of probes is separately  
 CC and addressably isolatable or amplifiable from the plurality), a single  
 CC exon microarray for measuring human gene expression, a method of  
 CC measuring human gene expression, a vector comprising the single exon  
 CC probe cited above, an ORF-encoded peptide comprising at least 8  
 CC contiguous amino acids of any of the above-mentioned amino acid  
 CC sequences (optionally with conservative amino acid substitutions), an  
 CC isolated antibody that binds specifically to a peptide cited above,  
 CC methods of selling and/or licensing single exon probes or microarrays to  
 CC a customer desiring to measure gene expression, a method of providing  
 CC human gene expression data by subscription, and a computer-readable  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above. The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterizing  
 CC alternative splicing events, in detecting and characterizing gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docid=2003194704  
 XX  
 SQ Sequence 994 BP; 252 A; 159 C; 203 G; 380 T; 0 U; 0 Other;  
 Query Match 7.4%; Score 92; DB 12; Length 994;  
 Best Local Similarity 52.0%; Pred. No. 1.2e-08;  
 Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;  
 QY 7 GAAAGTGAGAAACCAATGATATCATCTTTATATCTTCTTCTTATCATCGTCC 66  
 Db 763 GAATGTGGCAAGCTTCTTCAACCTTCTTCAACCTTCTTCAACATTAAGTATTTCACT 704  
 QY 67 AAAAGTATTTTGCACATATGAGGTTGTGAAGAGCCATTAATGACCATCATTTATTA 126  
 Db 703 GAGAGAGAGCCCTTCAATGTGAAGATGTGTAAGCTTTTAAACACCTTCAAGCCCTT 644  
 QY 127 GAGCAATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAAGATGTG 186  
 Db 643 ACTACACATTAAGTTCATTTATGTTAAAGAAAAACCTTCAAAATGT-----GAAAGATGT 590

```
OY 187 GATAAGCATTTTTCAGAAAATCACATTGGAAAACATATTGTATCATTCGAAAA 246
DB 589 GACAAAGCTTTTAAACGATTCATACCTTACTAAACATTAATTCATTCGAGAG 530
OY 247 AAACATTCATTTGTCAGTGTGGTAAAGGGGTAATTCTCGACAACCTGAAAAA 306
DB 529 AATCTTCAATGATGACATGTGGCAAAAGGCTTTAACTGGTCTTCAACCTTACAAA 470
OY 307 CATGAATACCCATACAAAGTCATTAAATGTACATTGGAAATTTGCAAGAGCATTT 366
DB 469 CATTAAGAATTCATCTGGAGAGAAACCTACAAATGTGAAGAAATGTGGCAAAAGCTTT 410
OY 367 TATTAACAT---CAATCTTAAGACATCATATATATCTGTTGATGAAAAAACATTACG 423
DB 409 AATGTCTCTTCACACCTTACTACACATTAAGATATTCATATCTGAGAGAAACCTTACAA 350
OY 424 TGTAAACATGTATTAAGTTTCTACTGACCTTCAAAATTTAGCAACATTAATTA 483
DB 349 TGTGAAGAAATGTGGCAAGGCTTTAAACCACTCTCAAACTTATCTATTAAGATTAAT 290
OY 484 CATCATGTGATCTCTCT 501
DB 289 CATACTGGAGAGAAACCT 272

RESULT 12
ADRI4368
ID ADRI4368 standard; DNA; 2873 BP.
AC ADRI4368;
XX
XX
XX 21-OCT-2004 (first entry)
DE Human NF-kappaB pathway-associated gene SeqIDJ369.
XX
XX NF-kappaB pathway; antiinflammatory; cytosolic; hepatotropic; virucide;
XX antiarthritic; antirheumatic; gastroinestinal-gen; antiasthmatic;
XX antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;
XX immunosuppressive; vulnereary; gene therapy; immune disorder;
XX inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;
XX hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;
XX hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;
XX X-linked anhidrotic ectodermal dysplasia; immunodeficiency;
XX viral infection; HIV-1; hepatitis B; hepatitis C; EBV; influenza;
XX rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;
XX atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;
XX autoimmune disorder; hyper immune activity;
XX aberrant acute phase response; hypercongenital condition; birth defect;
XX necrotic lesion; wound; organ transplant rejection;
XX aberrant signal transduction; proliferating disorder; cancer;
XX HIV propagation; gene; ds; human.
XX
XX Homo sapiens.
XX
XX WO2004065577-A2.
XX
XX 05-AUG-2004.
XX
XX 13-JAN-2004; 2004MO-US000798.
XX
XX 14-JAN-2003; 2003US-0440068P.
XX
XX 12-MAY-2003; 2003US-0469757P.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Nadler SG, Neubauer MG, Feder JN, Carman J;
XX
XX WPI; 2004-562168/54.
XX
XX P-PSDB; ADRI4369.
XX
XX New isolated polynucleotides and polypeptides associated with NF-kappaB
PT
```

PT pathway, useful for diagnosing, treating, or preventing disorders or  
PT diseases associated with NF-kappaB pathway.  
XX  
XX Claim 1, SEQ ID NO 369; 237bp; English.

CC This invention relates to the novel association of protein sequences (and  
CC the genes which encode them) to the NF-kappaB pathway. The invention may  
CC be useful for the production of compounds with an antiinflammatory,  
CC cyclostatic, hepatotropic, virocidic, antiarthritic, antirheumatic,  
CC gastroinestinal-gen, antiasthmatic, antiarteriosclerotic,  
CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or  
CC vulnereary activity or for gene therapy. The proteins and nucleotides are  
CC useful for diagnosing, preventing, treating, or ameliorating conditions  
CC or diseases associated with the NF-kappaB pathway. The condition is an  
CC immune disorder, an inflammatory disorder, an inflammatory disorder  
CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,  
CC hepatic disorder, Hodgkin's lymphoma, haematopoietic tumours, hyper-IgM  
CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic  
CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,  
CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell  
CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory  
CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick  
CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper  
CC immune activity, disorders related to aberrant acute phase responses,  
CC hypercongenital conditions, birth defects, necrotic lesions, wounds,  
CC organ transplant rejection, conditions related to organ transplant  
CC rejection, disorders related to aberrant signal transduction,  
CC proliferating disorders, cancers and HIV propagation in cells infected  
CC with other viruses. The present sequence is that of a human gene which is  
CC subject to the novel association with the NF-kappaB pathway of the  
CC invention. Note: This sequence does not appear in the specification but  
CC was obtained by the indexer from Genbank.  
XX  
XX

Seq Sequence 2873 BP; 948 A; 564 C; 545 G; 816 T; 0 U; 0 Other;

Query Match 74%; Score 92; DB 13; Length 2873;

Best Local Similarity 52.2%; Pred. No. 1.4e-08;

Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

```
OY 7 GAAATGACGAAACCAATTCATATCATCTTAAATATCTCTCTTCATCAGTCC 66
DB 1621 GAATGTCCGAAACATTTTTCATGCTTTCACACCTTAACATCAATTAAGATCCAACT 1680
OY 67 AAAAAGTATTTTGGACATATGAGGGTGTGTAAAGCCTATATATGACATATTATTA 126
DB 1681 AGAGGAATTTCTCAAAATGTGAAGCATATGGAAGGCTTTAACTGGTCCCAACCTT 1740
OY 127 GAGCAACATTTAAGAACCCAGTATATGATGACCGTATTAATGTACGTGACGATGT 186
DB 1741 AATTAACATTAAGAAATTCATATCTGAGAAAAAACCCTTCAATATGTAA-----GAATGT 1794
OY 187 GATAAGCATTTTTCAGAAAATCACATTGGAAAACATATTTGTATCATTCGAAAAA 246
DB 1795 GGCNAAGCCTTTAAACCAAGCTCACACTTATTAACATTAAGAAATTCATCTGAAG 1854
OY 247 AAACATTCATTTGTCAGTGTGGTAAAGGGGTTAAATTCGCAACACCTTGAAGAA 306
DB 1855 AAACCTTAATGATGAAAGATGTGGCAAAAGCCTTTAAACAGTATCGACCTTACTCA 1914
OY 307 CATGAATACCCATACAAAGTCATTAAATGTACATTGGAAATTTGCAAGAGCATTT 366
DB 1915 CATATATATTAATTCATATCTGGAGAAATTCCTTCAAAATGTGAAGAAATGTGTAGACCTTTT 1974
OY 367 TATTAACATCATCTTTAAGACATCATATATATATCTGT---CATGAAGAAACATTACG 423
DB 1975 AACCAAGCCTCAAGGCTTACTGACATTAAGTAAATTCATACCGGAGAGAAAGTTATGAA 2034
OY 424 TGTAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGCAACATTAATTA 483
DB 2035 TGTGAAGAAATGTGGCAAGGCTTTTAAACGATCTCAAAACTTACTGAACATTAAGATTA 2094
OY 484 CATCATGTG 493
```

Db 2095 CATACTGAG 2104

RESULT 13

ADN99064

ID ADN99064 standard; cDNA; 1533 BP.

XX

AC ADN99064;

XX

DT 29-JUL-2004 (first entry)

XX

DE Novel human cDNA sequence #664.

XX

XX de; gene; anti-inflammatory; dermatological; neuroprotective; immunomodulator; antibacterial; virucide; antiproliferative; cytostatic; gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes; early aging; hormonal imbalance; ischemic heart disease; ulcerative colitis.

XX

OS Homo sapiens.

XX

PN WO2004038003-A2.

XX

PD 06-MAY-2004.

XX

PF 24-OCT-2003; 2003MO-US033947.

XX

PR 25-OCT-2002; 2002US-0421061P.

PR 25-OCT-2002; 2002US-0421080P.

PR 25-OCT-2002; 2002US-0421552P.

PR 25-OCT-2002; 2002US-0421614P.

PR 30-OCT-2002; 2002US-0422177P.

PR 30-OCT-2002; 2002US-0422178P.

PR 15-NOV-2002; 2002US-0426355P.

PR 15-NOV-2002; 2002US-0426384P.

PR 15-NOV-2002; 2002US-0426394P.

PR 15-NOV-2002; 2002US-0426430P.

PR 15-NOV-2002; 2002US-0426916P.

PR 27-NOV-2002; 2002US-0429224P.

PR 27-NOV-2002; 2002US-0429275P.

PR 27-NOV-2002; 2002US-0429302P.

PR 27-NOV-2002; 2002US-0429326P.

PR 27-NOV-2002; 2002US-0429651P.

PR 04-DEC-2002; 2002US-0430645P.

PR 04-DEC-2002; 2002US-0430651P.

PR 04-DEC-2002; 2002US-0430653P.

PR 04-DEC-2002; 2002US-0430663P.

PR 04-DEC-2002; 2002US-0430684P.

PR 05-DEC-2002; 2002US-0430937P.

PR 05-DEC-2002; 2002US-0430965P.

PR 05-DEC-2002; 2002US-0431458P.

PR 12-DEC-2002; 2002US-0433251P.

PR 12-DEC-2002; 2002US-0433500P.

PR 13-DEC-2002; 2002US-0433316P.

PR 13-DEC-2002; 2002US-0433318P.

PR 23-DEC-2002; 2002US-0436236P.

PR 03-JAN-2003; 2003US-0437914P.

PR 17-JAN-2003; 2003US-0440820P.

PR 17-JAN-2003; 2003US-0440821P.

PR 18-APR-2003; 2003US-0463700P.

PR 18-APR-2003; 2003US-0463708P.

PR 18-APR-2003; 2003US-0463716P.

PR 18-APR-2003; 2003US-0463732P.

PR 02-MAY-2003; 2003US-0467199P.

PR 02-MAY-2003; 2003US-0467201P.

PR 02-MAY-2003; 2003US-0467203P.

PR 02-MAY-2003; 2003US-0467230P.

PR 19-MAY-2003; 2003US-0471306P.

PR 19-MAY-2003; 2003US-0471336P.

PR 22-MAY-2003; 2003US-0472420P.

PR 22-MAY-2003; 2003US-0472430P.

PR 09-JUN-2003; 2003US-0476609P.

PR 09-JUN-2003; 2003US-0476621P.

PR 09-JUN-2003; 2003US-0476632P.

PR 09-JUN-2003; 2003US-0476641P.

PR 08-JUL-2003; 2003US-0485217P.

PR 08-JUL-2003; 2003US-0485218P.

PR 08-JUL-2003; 2003US-0485223P.

PR 08-JUL-2003; 2003US-0485224P.

PR 08-JUL-2003; 2003US-0485252P.

PR 08-JUL-2003; 2003US-0485359P.

PR 14-JUL-2003; 2003US-0486446P.

PR 14-JUL-2003; 2003US-0486480P.

PR 15-JUL-2003; 2003US-0486891P.

PR 15-JUL-2003; 2003US-0486960P.

PR 08-AUG-2003; 2003US-0493341P.

PR 08-AUG-2003; 2003US-0493370P.

PR 08-AUG-2003; 2003US-0493573P.

PR 08-AUG-2003; 2003US-0493577P.

XX

XX (FIVE-) FIVE PRIME THERAPEUTICS INC.

XX

XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D,

PI Halendeck RF, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y,

PI Wong JGP, Wu G, Zhang H, Zeng C;

XX

DR WPI; 2004-365511/34.

DR P-PsDB; ADN99848.

XX

PT New nucleic acid molecules, useful in preparing a composition for

PT treating or preventing e.g. inflammatory, CNS, bacterial or viral

PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or

XX ulcerative colitis.

PS Claim 1, SEQ ID NO 664; 532pp; English.

XX

CC The invention relates to a nucleic acid molecule comprising a

CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or

CC preventing inflammatory, CNS, immune, bacterial or viral disorder,

CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic

CC heart disease or ulcerative colitis. This sequence corresponds to a

XX nucleic acid of the invention.

XX

SQ Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 12; Length 1533;

Best Local Similarity 52.1%; Pred. No. 1.4e-08;

Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

QY 11 GTGACGAAACCAATGATATATCATCTTTAATCTCTTCTTCATCACGCCCAAA 70

Db 305 GTGGCAATCATATTTGATGCTTTTCAATTTAACAATTAAGAAATTTCACTAGAG 364

QY 71 AGTATTTTGCACATATGAAGGATGATTAAGCTATATGACCATGATTTAGAGC 130

Db 365 AGTATTTTCAAAATGTAAGATGTAAGATGTAAGCTTTAACTGATGCTCAACCTTACAT 424

QY 131 AACATTAAAGAACCAACAGTATGATGACCGTATTAATGTAAGTGAAGATGTAAGT 190

Db 425 AACATTAAGATTAATTCATCTGAGAAACCCCTCAAAATGT-----GAAGATGTGGCA 478

QY 191 AAGCATTTTTCGAANAATCATCTTGGAAACATATGTAATGATCATTCGGAATAAAAC 250

Db 479 AAGCTTTTAACCGGCTCTCAAAATCTTAACATTAATAAATTAATTCATCTGAGAGAAAC 538

QY 251 CATTCATTTTCAAGTGTGTGTAAGGAGGATTAATCTCGACAAACCTGAAAGACATG 310

Db 539 CTTACAAATGTGAAGATGTGGAAAGCTTTTAACCGGCTCAACCTTAACTAAACAT 598

QY 311 AATCAACCATTAAGATCAATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 370

Db 599 AAGATTCATACAGAAAGAAACCTTACAAATGTAAGATGTGCAAGGCTTTAAC 658

QY 371 AACATCAATCTT---AAGACATCATATATATCTCTGTCATGAAATAAACATTAACTGTA 427



QY 251 CATTCATTGTCAGTGTGTAAAGGGGTAAATTCGACACACCTTGAAAAGACATG 310  
DB 539 CTTACAAATGTGAAGATGTGGCAAGCTTTAAACGGTCTCAACCTTACTAAACATA 598  
QY 311 AATTCACCATACAAAGTCATTTAAATGTACATTTGAAAATTTGTCAAGAACATTTTATA 370  
DB 599 AAGAAATTCATACAGAAAGAAACCTTACAAATGTAGAAATGTGCAAGGCTTTAAC 658  
QY 371 AACATCAATCTT---AAGACATCATATATATCTGTCTGAAAAAACATTAAGCTTA 427  
DB 659 AGTTCGATTTCTTAATTAACATTAAGAAATTCATATGAGATTAACCTTACAAATGTG 718  
QY 428 AACATGTATTAAGTTTCTACTCGACCTTCAAAATTTAGCAACAATTAATTAACATC 487  
DB 719 AAGAAATGTGCAAAAGCTTTAGAGTATTCCTCAATTTCTTAAAAACATTAAGATTAATCATA 778  
QY 488 ATGTGTGATCTCC 500  
DB 779 CTGGGAAAAAAC 791

RESULT 15  
ADB63624  
ID ADB63624 standard; cDNA; 2110 BP.  
XX  
AC ADB63624;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human cDNA encoding clone THYM20071120.  
XX  
KM Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
KM tissue regeneration; cell regeneration; membrane protein;  
KM signal transduction-related protein; transcription-related protein;  
KM osteoporosis; neurological disease; cancer; tumour.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 170..1792  
FT /tag= a  
FT /product= "Clone THYM20071120 protein"

PN BP1308459-A2.  
XX 07-MAY-2003.  
PD 28-MAR-2002; 2002EP-00007401.  
PF 05-NOV-2001; 2001JP-00379298.  
PR 25-JUN-2002; 2002US-00350578.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Iocga T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S,  
PI Yamamoto J, Isono Y, Hito Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y,  
XX  
DR WPI; 2003-450961/43.  
DR P-PSDB; ADB65594.  
XX  
PT New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
XX  
PS Claim 1; Page; 222pp; English.  
XX  
CC The invention discloses a polynucleotide comprising a sequence selected  
CC from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or

CC peptide of the polynucleotide by contacting the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesizing the polynucleotide, and as a probe  
CC for detecting the polynucleotide. The polynucleotides and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumour). The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.  
XX  
SQ Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 10; Length 2110;  
Best Local Similarity 51.7%; Pred. No. 1.5e-08;  
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

QY 61 CGTCCCAAAAGATATTGTCATATGAAAGGCTGATTAAGCCATATGACACATCA 120  
DB 1253 CATCCGAAGAGAAACCTTACAAATGTAAAGATGTGGCAAGCTTTTAAACCTCTCA 1312  
QY 121 TTATTAGCAACATTTAAGAAACCCACAGTAATGATCGACCGTATTAATGTACGTGAC 180  
DB 1313 GCCCTTACACATTAAGAAATTCACACTGAGAGAAACCCACAAATGT-----GAA 1366  
QY 181 GATTGTGATTAAGCAATTTTTCAGAAATTCACATTTGGAACACATATTTGATACATTC 240  
DB 1367 GAATGTGCAAAAGCTTTTAAACGATCTCAAAACCTTACGAACTAAGAAACCTTACTACT 1426  
QY 241 GAAAAAAACCAATTCATGTTCAAGTGTGTAAGGGGTAAATCTGCAACAACCTTG 300  
DB 1427 GGAAGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTACCAATCTTCAAACTT 1486  
QY 301 AAAAGACATGAATCACCCATCAAAAGTCATTTAATGTACATTTGAAAATTTGTCAAGAA 360  
DB 1487 ACTGAACATTAATAAATTTCTTGTGAGAGATACCTTACAAAGTGAAGATGTGGCAAA 1546  
QY 361 GCATTTTAAACATCAATCTTTA---AGACATCATATATATCTGTCATGAAAAACA 417  
DB 1547 GCTTTTAAACACTCTCATCCCTTACTACACATTAAGAAATTCATCTGGGGAACCC 1606  
QY 418 TTACGTGTAACAAATGTATTAAGTTTCACTCCGACCTTCAAAATTTAGCAACAATTA 477  
DB 1607 TACAAATGTGAAGATGTGCAAAAGCTTTTAAAGCCGATCTTCAAACTTACTGAACAT 1666  
QY 478 TTAAACATCATGCTG-TGATCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAA 536  
DB 1667 ATPATTCATACCTGAGAGAAACCTTAAATGTGAAGATGTGACAAAGCTTTTAAACCA 1726  
QY 537 TTTCAAACTGTGTCAGTATTAATTTCAATATTAATAAACAATGATCAAAACTTAAATG 596  
DB 1727 TCTGCAAACT-----TCTAAACATTAATAAATAATCATATCTGAGAGAAATCAAGAAC 1780  
QY 597 TCCTAAATGTGTAAGAGTTGTGTGGAAAAAGTTTATCTTCAATATGTAAATGTC 656  
DB 1781 TGGAAATGTGATTAAGATTTTGAACAACCTTCAAAATTTTCTTAATTAAGAAATCA 1840  
QY 657 TGATGAT 663  
DB 1841 TACTGAT 1847

Search completed: February 9, 2005, 02:39:51  
Job time : 644.278 secs



This Page Blank (uspto)



QY 976 GCTTCTTGGAGAGTTCAGAAAGAGTCTTCTTAACTTATCTGATATAGTGGAGAGAGATC 1035  
| | | | |  
Db 364 GCTTCTTGGAGAGTTCAGAAAGAGTCTTCTTAACTTATCTGATATAGTGGAGAGAGATC 423  
| | | | |  
QY 1036 AATGTCCTAGAGATTAATGTGTAGTAAGTCTTCTAGAGATATGATTTAAGTGCACAT 1095  
| | | | |  
Db 424 AATGTCCTAGAGATTAATGTGTAGTAAGTCTTCTAGAGATATGATTTAAGTGCACAT 483  
| | | | |  
QY 1096 TTGAATGGCATGATATATTTACAAAGATTTGAGTCACTTCTTAATAGTATGAAAAA 1155  
| | | | |  
Db 484 TTGAATGGCATGATATATTTACAAAGATTTGAGTCACTTCTTAATAGTATGAAAAA 543  
| | | | |  
QY 1156 GAAGAACTCCAGAGAGTGAACCATTTGGTTAAAAAAGCCAGATGATTTATGCCAAAT 1215  
| | | | |  
Db 544 GAAGAACTCCAGAGAGTGAACCATTTGGTTAAAAAAGCCAGATGATTTATGCCAAAT 603  
| | | | |  
QY 1216 GAAACATCAGTATTTCTGATTA 1239  
| | | | |  
Db 604 GAAACATCAGTATTTCTGATTA 627  
| | | | |

## RESULT 2

US-09-248-796A-5186  
; Sequence 5186, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248, 796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074, 725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096, 409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 5186  
; LENGTH: 462  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-5186

Query Match 31.6%; Score 391; DB 4; Length 462;  
Best Local Similarity 98.5%; Pred. No. 4.4e-80;  
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGAGTGAAGTGAAGCAACCAATGATATCATCTTTAATCTTCTTCTTCATCA 60  
| | | | |  
Db 52 ATGAGTGAAGTGAAGCAACCAATGATATCATCTTTAATCTTCTTCTTCATCA 111  
| | | | |  
QY 61 CGTCCCAAAAAGATTTTTCATATGAGAGGTGTGAAGGCCATATATGACCATCA 120  
| | | | |  
Db 112 CGTCCCAAAAAGATTTTTCATATGAGAGGTGTGAAGGCCATATATGACCATCA 171  
| | | | |  
QY 121 TTATTAGCAACATTTAAGAAACCAAGTAAATGATGACCGTATTAATGTCAGTGAC 180  
| | | | |  
Db 172 TTATTAGCAACATTTAAGAAACCAAGTAAATGATGACCGTATTAATGTCAGTGAC 231  
| | | | |  
QY 181 GATTTGATTAAGCATTTTTCAGAAAAATCACTTTGAAAAACATATGATATCATCTTC 240  
| | | | |  
Db 232 GATTTGATTAAGCATTTTTCAGAAAAATCACTTTGAAAAACATATGATATCATCTTC 291  
| | | | |  
QY 241 GAAAAAACCATTCCATGTTCAAGTGTGTGAAGGGGTAAATCTGCAACAACCTTG 300  
| | | | |  
Db 292 GAAAAAACCATTCCATGTTCAAGTGTGTGAAGGGGTAAATCTGCAACAACCTTG 351  
| | | | |  
QY 301 AAAAGCATGAATCAACCATACAAAGTCAATTTAAATGATGATTTGAAATTTGCAAGA 360  
| | | | |  
Db 352 AAAAGCATGAATCAACCATACAAAGTCAATTTAAATGATGATTTGAAATTTGCAAGA 411  
| | | | |  
QY 361 GCATTTAT-AAACATCAATCTTTAAGACATATATATCTGTCATGA 410  
| | | | |

Db 412 ACATTTTAAAAACATCAATCTTTAAGACATATATATATCTGTCATGA 462  
| | | | |

## RESULT 3

US-09-016-434-1054  
; Sequence 1054, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016, 434  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1054:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2320 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g1017721  
US-09-016-434-1054

Query Match 7.8%; Score 96.6; DB 4; Length 2320;  
Best Local Similarity 51.3%; Pred. No. 1.4e-12;  
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGATATTTTTCATATGAGAGGTGTGATTAAGCCATATATGACCATCATTTATGA 128  
| | | | |  
Db 964 AAGAAACCCCTACCAATGTAAAGATGTGTAAAGCTTTTAACGATCTTGACCCCTAC 1023  
| | | | |  
QY 129 GCACATTTAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGGAGATGTGA 188  
| | | | |  
Db 1024 TACCATGAAAAAATTTCACTGAGAGAAACCTTACAAATGT-----GAAAGATGTGG 1077  
| | | | |  
QY 189 TAAACATTTTTCAGAAAAATCACTTTGAAAAACATATGATATCACTTCGAAAAAAA 248  
| | | | |  
Db 1078 CAAACCTTTTAAAGCAGTCTTCAAACTTACATACATTAAGATTAATCTACTGAGAGAA 1137  
| | | | |  
QY 249 ACCATTCATGTTCAAGTGTGTGAAGGGGTAAATCTGCAACAACCTTGAAGAAGCA 308  
| | | | |  
Db 1138 ACCCTACAAATGTAAAAATGTGAAGGCCCTTTAACACAGTCTGACACCTTACACACA 1197  
| | | | |  
QY 309 TGAATCAACCATACAAAGTCAATTTAAATGTATCAATTTGAAAAATGTCAAGAGATTGA 368  
| | | | |

```
Db 1198 TGAAGTAATTCATCTGAGAGAAACCTCAAAATGTGAAAAATGTGAAAAACCTTTAA 1257
Qy 369 TAAACA---TCATCTTTAAGACATATATATGTTTCATGAAAAAATTAACGCG 425
Db 1258 TCATTTCTACACCTTACTACATAGAAATTCATACGTGAGAGAAACCTTACAAAG 1317
Qy 426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGCAACATTAATTAAGA 485
Db 1318 TAAAGATGTGTAAGCTTTTAAACATCTTCAACCTTACAAACATTAAGTAATTCG 1377
Qy 486 TCATGTGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAAAATTTCCAAAC 545
Db 1378 TACTGGAGAGAAAGCTTACAAATCTTAAGATGTGAAAAAGCTTTTAAACCAATCTCAA 1437
Qy 546 TTGCTGAGTATTAATTTATTAATAACATGTCATCCAAAATTAATGTCCTTAAG 605
Db 1438 ACTTACTGAACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTGAAAAAG 1494
Qy 606 TGGTAAGGTGTGTTGGGAAAAAGTTTATCTTGCATATGTTAAGTCATGATGA 662
Db 1495 TGGCAAGCTTTTAAACCATGCTCAATCTTACTAAGCATTAAGAAAGTCATACAGA 1551

RESULT 4
US-09-949-016-4204
; Sequence 4204, Application US/09949016
; Patent No. 6812338
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4204
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4204

Query Match 7.3%; Score 90.8; DB 4; Length 3798;
Best Local Similarity 53.9%; Pred. No. 3.4e-11;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

Qy 69 AAAGTATATTTGCATATGAAAGGTGTGATTAAGCCTTAATTCGACCATTAATTAGA 128
Db 1848 AGGAAACCCCTACAAATGTGAAGATGTGCAAAAGCTTTAGCATTTCTCAACCTTGC 1907
Qy 129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTAAATGTGACAGTGAGATTGTA 188
Db 1908 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAAATGTG 1961
Qy 189 TAAAGCATTTTTCAGAAATTCATTTTGAAACATATGTATGTCATTTCCGAAAAAAA 248
Db 1962 CAAAGCTTTTACGCGTTCTTCAACCTTGTAAACATTAAGAAATTCATCTGAGAGAA 2021
Qy 249 ACCATTTCATTTGTAGTGTGTGTAAGAGGTTAAATTCGCAACACCTTGAAGAGACA 308
Db 2022 ACCCTCAATGTAAAGAAATGTGCAAAAGCTTTTGAATTCCTCAACCTTGCTATACA 2081
Qy 309 TGAATTCACCCATCAAAAGTCATTTAATGTACATTTGAAATTTGCAAGAGCATTTTA 368
Db 2082 TAAGATTAATCTCATCTGAGAGAAACCTTCAAAATGTAAAGAAATGTGACAAACCTTTTA 2141
Qy 369 TAAACATCATC---TTTAAAGATCATATATATCTGTTCAAGAAAAACATTAACGCG 425
```

```
Db 2142 GGCATCTCAACCCCTTACTAAACATTAATTAATACATGCTGAGAGAAACCTTACAAAG 2201
Qy 426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGCAACATTAATTAAGA 485
Db 2202 TGAAGATGTGCAAAAGCTTTTAAATGTGATCTTCAATCTTACTATTAACATGTTTATGA 2261
Qy 486 TCATGTGATCTCTCC 501
Db 2262 TACTGGAGAGAAACCT 2277

RESULT 5
US-09-949-016-12227
; Sequence 12227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12227
; LENGTH: 156942
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(156942)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12227

Query Match 7.3%; Score 90.8; DB 4; Length 156942;
Best Local Similarity 53.9%; Pred. No. 1.1e-10;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

Qy 69 AAAGTATATTTGCATATGAAAGGTGTGATTAAGCCTTAATTCGACCATTAATTAGA 128
Db 152999 AGGAAACCCCTACAAATGTGAAGATGTGCAAAAGCTTTAGCATTTCTCAACCTTGC 153058
Qy 129 GCAACATTTAAGAACCCACAGTAATGATGACCGCTTAATGTACAGTGAGATTGTA 188
Db 153059 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAAATGTG 153112
Qy 189 TAAAGCATTTTTCAGAAATTCATTTTGAAACATATGTATGTCATTTCCGAAAAAAA 248
Db 153113 CAAAGCTTTTACGCGTTCTTCAACCTTGTAAACATTAAGAAATTCATCTGAGAGAA 153172
Qy 249 ACCATTTCATTTGTAGTGTGTGTAAGAGGTTAAATTCGCAACACCTTGAAGAGACA 308
Db 153173 ACCCTCAATGTAAAGAAATGTGCAAAAGCTTTTGAATTCCTCAACCTTGCTATACA 153232
Qy 309 TGAATTCACCCATCAAAAGTCATTTAATGTACATTTGAAATTTGCAAGAGCATTTTA 368
Db 153233 TAAAGTAACTCATCTGAGAGAAACCTTCAAAATGTAAAGAAATGTGACAAACCTTTTA 153292
Qy 369 TAAACATCATC---TTTAAAGATCATATATATCTGTTCAAGAAAAACATTAACGCG 425
Db 153293 GGCATCTTCAACCCCTTACTAAACATTAATTAATACATGCTGAGAGAAACCTTACAAAG 153352
Qy 426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGCAACATTAATTAAGA 485
Db 153353 TGAAGATGTGCAAAAGCTTTTAAATGTGATCTTCAATCTTACTATCATTAAGTTTATGA 153412
```

QY 486 TCATGTGATCTCCT 501  
DB 153413 TACTGGAGAGAAACCT 153428

## RESULT 5

US-09-949-016-15946  
Sequence 15946, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15946  
LENGTH: 156950  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(156950)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15946

Query Match  
Best Local Similarity 53.3%; Score 90.8; DB 4; Length 156950;  
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAGGGGTGATTAAGGCTTAATTCGACATCTATTATTA 128  
DB 153999 AGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTAGCATTTCTCAACCTTGC 153058  
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTAACAGTGA 188  
DB 153059 TAAACATAGAGAAATCATACGTGAGAGAAACCTCAATGT-----GAAGATGTG 153112  
QY 189 TAAAGATTTTCAAGAAATCAATTGGAAACATATTGTAATCATTTCCGAAAAA 248  
DB 153113 CAAAGCTTTAGCGCTTCTTCAACCTTGTCTAAACATTAAGAAATTCATACGTGAGAGA 153172  
QY 249 ACCATTCCATTGTTCAAGTGTGTAAAGGGTTAATTCGCAACAACCTGAAAGACA 308  
DB 153173 ACCCTCAATGTGAAGATGTGGCAAGCTTTAGCAATTCCTCAACCTTGTCTATCA 153232  
QY 309 TGAATACCCATACAAAGTCAATTAATGTACATTGAAAAATGTCAAGAAACATTTTA 368  
DB 153233 TAAGATTAATCTAATGAGAGAAACCTTCAAAATGAATGAAGATGACAAACCTTTTA 153292  
QY 369 TAAACATCAATC---TTTAAGACATATATATATCTGTCATGAAAAAATTAACGTG 425  
DB 153293 GCGACTCTCAACCTTACTTAATCAATTAATTAATCTGAGAGAAACCTCAAAATG 153352  
QY 426 TAAACATGTAAATTAAGTTTCACTGACCTCAAAATTAAGCAACATTAATTAACA 485  
DB 153353 TGAAGATGTGGCAAGCTTTTATGATCTTCAAAATCTTACTATACATTAATTAACA 153412  
QY 486 TCATGTGATCTCCT 501  
DB 153413 TACTGGAGAGAAACCT 153428

RESULT 7  
US-09-949-016-485

/ Sequence 485, Application US/09949016  
/ Patent No. 6812339  
/ GENERAL INFORMATION:  
/ APPLICANT: VENTER, J. Craig et al.  
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
/ FILE REFERENCE: CL001307  
/ CURRENT FILING DATE: 2000-04-14  
/ PRIOR APPLICATION NUMBER: 60/241,755  
/ PRIOR FILING DATE: 2000-10-20  
/ PRIOR APPLICATION NUMBER: 60/237,768  
/ PRIOR FILING DATE: 2000-10-03  
/ PRIOR APPLICATION NUMBER: 60/231,498  
/ PRIOR FILING DATE: 2000-09-08  
/ NUMBER OF SEQ ID NOS: 207012  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 485  
/ LENGTH: 3839  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-949-016-485

Query Match  
Best Local Similarity 53.7%; Score 89.2; DB 4; Length 3839;  
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAGGGGTGATTAAGGCTTAATTCGACATCTATTATTA 128  
DB 964 AGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTAGCATTTCTCAACCTTGC 1023  
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTAACAGTGA 188  
DB 1024 TAAACATAGAGAAATTCATACGTGAGAGAAACCTCAATGT-----GAAGATGTG 1077  
QY 189 TAAACATTTTCAAGAAATCAATTGGAAACATATTGTAATCATTTCCGAAAAA 248  
DB 1078 CAAAGCTTTAGCGCTTCTTCAACCTTGTCTAAACATTAAGAAATTCATACGTGAGAGA 1137  
QY 249 ACCATTCCATTGTTCAAGTGTGTAAAGGGTTAATTCGCAACAACCTGAAAGACA 308  
DB 1138 ACCCTCAATGTGAAGATGTGGCAAGCTTTTAGCAATTCCTCAACCTTGTCTATCA 1197  
QY 309 TGAATACCCATACAAAGTCAATTAATGTACATTGAAAAATGTCAAGAAACATTTTA 368  
DB 1198 TAAGATTAATCTAATGAGAGAAACCTTCAAAATGAATGACAAACCTTTTA 1257  
QY 369 TAAACATCAATC---TTTAAGACATATATATATCTGTCATGAAAAAATTAACGTG 425  
DB 1258 GCGACTCTCAACCTTACTTAACATTAATTAATCATCTGAGAGAAACCTTCAAAATG 1317  
QY 426 TAAACATGTAAATTAAGTTTCACTGACCTCAAAATTAAGCAACATTAATTAACA 485  
DB 1318 TGAAGATGTGGCAAGCTTTTATGATCTTCAAAATCTTACTATACATTAATTAACA 1377  
QY 486 TCATGTGATCTCCT 501  
DB 1378 TACTGGAGAGAAACCT 1393

RESULT 8  
US-09-949-016-27359/c  
Sequence 27359, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20

```
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 27359
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-27359
```

```
Query Match
Best Local Similarity 51.4%; Pred. No. 1.6e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;
```

```
7.0%; Score 86.8; DB 4; Length 601;
7
GAAAGTGACGAAACCAATGATATCATCTTAAATCTTCTTCTTCATCAGTCCC 66
538 GAATGTGGCAAAAGCTTTTAAAGCACTCCCTGCTTAAACATTAATATCATGTCT 479
67 AAAAGTATATTGGCATATATGAAAGGTGTGATTAAGCCCTATATGACATCATATTA 126
478 GAGAGAAACTTACAAATGTGAGAAATGTGCAAAAGCTTTATCATCTTCAAAATCTT 419
127 GAGCAACATTTAAGAAACCCAGATATGATCGACCGTATTAATGTACAGTGAAGATGT 186
418 AGCACATTAAGTAAATTTCTACTTAAGAAACCTTCCAGAGT-----GAAAGATGT 365
187 GATTAAGCATTTTTCAGAAAATCATTGGAACAATTTGTATCATCTCCGAAAAA 246
364 GACAAAGCATTTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAG 305
247 AAACCATTCATTTGTCAGTGTGTGTGAAGGGGTTAATCTGTGACAAACCTTGAAGA 306
304 AAACCTTCAAAATGTGAAAGATGTGCAAAAGCATTTAGCCAGCTTCAACCTTACTACA 245
307 CATGAATCACCCATACAAAGTCAATTTAAATGTACATTTGAAATTTGCAAGAAAGATT 366
244 CATTAAGAGATGACACTGAGAGAAACCTTACAAATGTGAAGAAATGTGCAAAAGCTTT 185
367 TATTAACATCATATC---TTTAAGACATCATATTAATCTGTTCATGAAAAAATTAAG 423
184 AGCCAAATCTCAACCTTACTACATCAATTAATTAATTAATTAATTAATTAATTAATTA 125
424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAATTA 483
124 TGTGAAGAAATGTGCAAAAGCTTTTGAAGAAATCTTCACTGAAATTAAGATTAAT 65
484 CATCATGTGTG 493
64 CATACTGGAG 55
DB
```

```
RESULT 9
US-09-949-016-149638/C
/ Sequence 149638, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 149638
```

```
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-149638
```

```
Query Match
Best Local Similarity 51.4%; Pred. No. 1.6e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;
```

```
7.0%; Score 86.8; DB 4; Length 601;
7
GAAAGTGACGAAACCAATGATATCATCTTAAATCTTCTTCTTCATCAGTCCC 66
538 GAATGTGGCAAAAGCTTTTAAAGCACTCCCTGCTTAAACATTAATATCATGTCT 479
67 AAAAGTATATTGGCATATATGAAAGGTGTGATTAAGCCCTATATGACATCATATTA 126
478 GAGAGAAACTTACAAATGTGAGAAATGTGCAAAAGCTTTTATCATCTTCAAAATCTT 419
127 GAGCAACATTTAAGAAACCCAGATATGATCGACCGTATTAATGTACAGTGAAGATGT 186
418 AGCACATTAAGTAAATTTCTACTTAAGAAACCTTCCAGAGT-----GAAAGATGT 365
187 GATTAAGCATTTTTCAGAAAATCATTGGAACAATTTGTATCATCTCCGAAAAA 246
364 GACAAAGCATTTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAG 305
247 AAACCATTCATTTGTCAGTGTGTGTGAAGGGGTTAATCTGTGACAAACCTTGAAGA 306
304 AAACCTTCAAAATGTGAAAGATGTGCAAAAGCATTTAGCCAGCTTCAACCTTACTACA 245
307 CATGAATCACCCATACAAAGTCAATTTAAATGTACATTTGAAATTTGCAAGAAAGATT 366
244 CATTAAGAGATGACACTGAGAGAAACCTTCAAAATGTGAAGAAATGTGCAAAAGCTTT 185
367 TATTAACATCATATC---TTTAAGACATCATATTAATCTGTTCATGAAAAAATTAAG 423
184 AGCCAAATCTCAACCTTACTACATCAATTAATTAATTAATTAATTAATTAATTAATTA 125
424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAATTA 483
124 TGTGAAGAAATGTGCAAAAGCTTTTGAAGAAATCTTCACTGAAATTAAGATTAAT 65
484 CATCATGTGTG 493
64 CATACTGGAG 55
DB
```

```
RESULT 10
US-09-949-016-44799
/ Sequence 44799, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 44799
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-44799
```

```
Query Match
Best Local Similarity 54.4%; Pred. No. 2.2e-10;
7.0%; Score 86.2; DB 4; Length 601;
```



QY 247 AAACATTCATTGTCAGTGTGTGTAAAGGGTTAATTCGACAAACATTGAAAGA 306  
DB 265 AAACCTCAAAATGTGAAGAAATGTGGCAAGCAATTGACAGCCTTACACTACCA 206  
QY 307 CATGAATCCCAATACAAAGTCAATTTAAATGTACATTTGAAATTTGTCAAGACATTT 366  
DB 205 CATGAAGAGTGTGCAACTGTGAGAGAAACCTACAAATGTGAGAAATGTGCAAGCTTTT 146  
QY 367 TATAAATCATCAATC---TTTAAGACATATATATATCTGTGATGAAAAAAGATTAAAG 423  
DB 145 AGCCAAATCCTCAACCTTACTACACATTAATATTCATCTGAGAGAAACCTTACAA 86  
QY 424 TGTAAACAATGTAAATTAAGTTTCTACTGACCTTCAAAATTAACACATAATTTAAA 483  
DB 85 TGTGAAGAATGTGGCAAGCTTTTGAAGAAATCTTCAACTTACTGAACTAATGATTAAT 26  
QY 484 CATCATGTGTG 493  
DB 25 CATACTGGAG 16

RESULT 13  
US-09-949-016-149637/c  
; Sequence 149637, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 149637  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-149637

Query Match  
Best Local Similarity 51.4%; Pred. No. 3e-10;  
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAAGTGAGCAACCAATGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66  
DB 499 GAATGTGGCAAAAGCTTTTAAGACCTCCTCAGCCCTGTAAACATTAATAATACATGCT 440  
QY 67 AAAAGTATATTTGCAATATGAAAGGTGATTAAGCCATTAATGACACCATTAATTA 126  
DB 439 GGAAGAACTCTCAAAATGTGAGAAATGTGGCAAGCTTTTAATCAATCTTAAATCTT 380  
QY 127 GAGCAATTTAAGAACCCACAGTAATGATCGACCGTAAATGTACAGTGTGCAATGT 186  
DB 379 AGACACATTAAGTAATTTACTAATAAGAAACCTTCCAGAGT-----GAAGATGT 326  
QY 187 GATTAAGCATTTTTCAGAAAAATCACTTTGAAAAACATATTGTATCACTTCGAAAA 246  
DB 325 GACAAAGCATTTATCTGTCTCTCARCCCTTACGAACTAAGAGAAATTCATCCAGAG 266  
QY 247 AAACATTCATTGTCAGTGTGTGTAAAGGGTTAATTCGACAAACATTGAAAGA 306  
DB 265 AAACCTCAAAATGTGAAGAAATGTGGCAAGCAATTGACAGCCTTACACTACCA 206  
QY 307 CATGAATCCCAATACAAAGTCAATTTAAATGTACATTTGAAATTTGTCAAGACATTT 366  
DB 205 CATGAAGAGTGTGCAACTGTGAGAGAAACCTACAAATGTGAGAAATGTGCAAGCTTTT 146

QY 367 TATAAATCATCAATC---TTTAAGACATATATATATCTGTGATGAAAAAAGATTAAAG 423  
DB 145 AGCCAAATCCTCAACCTTACTACACATTAATATTCATCTGAGAGAAACCTTACAA 86  
QY 424 TGTAAACAATGTAAATTAAGTTTCTACTGACCTTCAAAATTAACACATAATTTAAA 483  
DB 85 TGTGAAGAATGTGGCAAGCTTTTGAAGAAATCTTCAACTTACTGAACTAATGATTAAT 26  
QY 484 CATCATGTGTG 493  
DB 25 CATACTGGAG 16

RESULT 14  
US-09-949-016-27361/c  
; Sequence 27361, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 27361  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-27361

Query Match  
Best Local Similarity 52.5%; Pred. No. 1e-09;  
Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

QY 57 ATCAGCTCCCAAAAGTATATTTGCAATATGAAGGGTGTGATTAAGCCTATATATGACC 116  
DB 554 AACTCATCTGAGAGAAACCTTACAAATGTAAAGATGTGCAAAACCTTTAAGGAGCT 495  
QY 117 ATCATATATGAGCAACATTTAAGAACCGACAGTATGATCGACGATTAATGATACAGT 176  
DB 494 CTCAACCTTACTAATCAATTAATTAATACATGCTGTGAGAGAACTTACAAATGT----- 440  
QY 177 GAGCAGTGTGTAAAGCAATTTTTCAGAAATCACTTTGAAACACATATGTATACCA 236  
DB 439 -GAAGAAATGTGGAAGAACTTTTAATGATCTTCAATCTTACTATCAATAGTTTATTTCA 381  
QY 237 TTCGAAAAAAACCATTCATTTTCAAGTGTGTGTAAGGGGTTAATTTCTGACACA 296  
DB 380 TACTGAGAGAAACCTTCAAGGTGAAAGATGTGCAAGCAATTAACGTGCTCTCAAG 321  
QY 297 CTTGAAAAACATGAAATCAACCATCAAGATCAATTAATGATCACTTTGAAATTTGTCA 356  
DB 320 CCTTACTAATCAATTAAGAAATTAATCTAGAGAGAAACCTTCAAAATGTAAGAAATGTG 261  
QY 357 AGAAGCATTTTATTAACATCAATCTTTA---AGACATCATATATTAATCTGTTCATGAAA 413  
DB 260 CAAGATCTTATATGTCTTCAACCTTACTGACATTAAGAGATCAACTGTGAGAGAA 201  
QY 414 AACATTAACGTGTAAACAATGTAATTAAGTTTCACTGACCTTCAAAATTAACACACA 473  
DB 200 GCCCTCAAAATGTGAAGATGTGGCAAGCTTTTACCGTCTTCAACCTTACTAAGCA 141  
QY 474 TAAATTAACAATCATGTGTG 493



Db 140 TAAGACAATTCATACTGGAG 121

```

RESULT 15
US-09-949-016-149636/c
; Sequence 149636; Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149636
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-149636

```

Query Match	6.7%;	Score 83.2;	DB 4;	Length 601;
Best Local Similarity	52.5%;	Pred. No. 1e-09;		
Matches 231;	Conservative 1;	Mismatches 199;	Indels 9;	Gaps 2;

Qy	5	ATACGCTCCCAAAAAGTATATATTGGA	CATATGAAGGGGTGATTAAGCTATATGACC	116
Db	554	AATCTACTGAGAGAAACCTTACAAATGTAAGAAATGTGACAAA	CTTTTAAGCACT	495
Qy	117	ATCATTTATGACCACTTATTAAGAACCCACAGTAATGATGCACCGTAAATGTACAGT		176
Db	494	CTCAACCCCTTACTTAAACATATAAATATATCATGCTGGAGAGAACTTACAAATGT	-----	440
Qy	177	GGACGATTTTGATTAAGCATTTTTCAGAAAATCATTTGGAAACACATATGTATCACA		226
Db	439	-GAAGATGTGGCAAAAGCTTTTAATGATCTTCAAAATTTATCTTACATTAAGTTTATTC		381
Qy	237	TTCCGAAAAAAAACCATTTGCATTTGTTCAGTGTGTGTAAGGGGTATATCTGCACACA		296
Db	380	TACTGAGAGAAACCTTACAATGTGAGAAATGTGGCAAAACATTTAACTGTGCTCCAG		321
Qy	297	CTTGAAAAAGCATGAATACCCCATACAAAGTCATTTAATGTACATTTGAAAATTGTCA		356
Db	320	CCCTTACTTAAACATTAAGAAATTCATACATGAGAGAAACCTTCAAATGTAAAGAAATGTGG		261
Qy	357	AGAGAGATTTTATTAACATCAATCTTTA---AGACATCATATATATCTGTTCATGAAA		413
Db	260	CANAGCATTTATATGTCTTCAACCTTAACCTTGAATATGAAGATACACACTGGAGGAA		201
Qy	414	AACATTAAAGTGTAAACATGTATATTAAGTTTCACTGCACCTTCAAATTTAGCACACA		473
Db	200	GCCTTCAATGTGAAAGATGTGGCAAGCCTTTTACCGCTTCTCAACCTTACTAAGCA		141
Qy	474	TAAATTTAAACATCATGTGTG		493
Db	140	TAAAGCAATTCATCTAGGAG		121

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 05:36:40 ; Search time 2160.11 Seconds  
(without alignments)  
3301.382 Million cell updates/sec

Title: US-09-831-804-2

Perfect score: 1239  
Sequence: 1 atgagtgaaagtcagcaaac.....catcagtcattctcgataa 1239

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4313806 seqs, 2877871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PC7\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PC7US\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
19: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
20: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1235.8	99.7	1239	US-10-032-585-6493	Sequence 6493, App
2	99.4	8.0	2132	US-10-094-749-726	Sequence 726, App
3	96.6	7.8	2320	US-10-305-720-1054	Sequence 1054, App
4	92	7.4	994	US-10-029-386-20763	Sequence 20763, App
5	92	7.4	2873	US-10-755-889-369	Sequence 369, App
6	91.8	7.4	2110	US-10-104-047-1778	Sequence 1778, App
7	91.8	7.4	3078	US-10-104-047-622	Sequence 622, App
8	90.4	7.3	2114	US-10-094-749-692	Sequence 692, App
9	90.2	7.3	2230	US-10-094-749-499	Sequence 499, App
10	89.8	7.2	590	US-10-029-386-13298	Sequence 13298, App
11	89.2	7.2	2221	US-10-108-260A-29	Sequence 29, App1

12	89.2	7.2	3839	US-10-172-118-910	Sequence 910, App
13	89.2	7.2	3839	US-10-342-887-910	Sequence 910, App
14	89	7.2	3502	US-10-098-841-48	Sequence 48, App1
15	88.6	7.2	1466	US-10-029-386-20574	Sequence 20574, App
16	88.4	7.1	1410	US-10-029-386-20532	Sequence 20532, App
17	88.4	7.1	2509	US-10-381-371-11	Sequence 11, App1
18	88.4	7.1	4132	US-10-198-846-13133	Sequence 13133, App
19	87.6	7.1	583	US-10-029-386-13093	Sequence 13093, App
20	87.6	7.1	1757	US-10-203-058B-6	Sequence 6, App1
21	87.2	7.0	831	US-10-264-049-975	Sequence 975, App1
22	86.8	7.0	2662	US-10-363-618-66	Sequence 66, App1
23	86.2	7.0	911	US-10-029-386-22636	Sequence 22636, App
24	86.2	7.0	1422	US-10-029-386-20322	Sequence 20322, App
25	86.2	7.0	1725	US-10-029-386-20556	Sequence 20556, App
26	86.2	7.0	1727	US-10-029-386-20367	Sequence 20367, App
27	86.2	7.0	2558	US-09-764-864-1664	Sequence 1664, App
28	86.2	7.0	2558	US-09-764-864-1665	Sequence 1665, App
29	86	6.9	2664	US-10-108-260A-511	Sequence 511, App
30	86	6.9	2690	US-10-723-860-8039	Sequence 8039, App
31	86	6.9	3400	US-10-104-047-1037	Sequence 1037, App
32	85.4	6.9	2505	US-10-108-260A-808	Sequence 808, App
33	84.8	6.8	2239	US-10-094-749-1084	Sequence 1084, App
34	84.4	6.8	1191	US-10-029-386-22992	Sequence 22992, App1
35	84.4	6.8	3138	US-10-104-047-83	Sequence 83, App1
36	84.2	6.8	1578	US-10-029-386-20214	Sequence 20214, App
37	84	6.8	565	US-10-029-386-13970	Sequence 3970, App
38	84	6.8	1619	US-10-422-522-46	Sequence 46, App1
39	84	6.8	3617	US-10-029-386-22868	Sequence 22868, App
40	84	6.8	5820	US-10-363-616-44	Sequence 44, App1
41	83.6	6.7	596	US-10-029-386-6664	Sequence 6664, App
42	83.2	6.7	1201	US-10-029-386-25405	Sequence 25405, App
43	83.2	6.7	2241	US-10-108-260A-1289	Sequence 1289, App
44	83.2	6.7	2525	US-10-108-260A-1358	Sequence 1358, App
45	83	6.7	500	US-10-029-386-6818	Sequence 6818, App

## ALIGNMENTS

RESULT 1					
US-10-032-585-6493					
; Sequence 6493, Application US/10032585					
; Publication No. US20030180953A1					
; GENERAL INFORMATION:					
; APPLICANT: Terry, Roemer D.					
; APPLICANT: Bo, Jiang					
; APPLICANT: Charles, Boone					
; APPLICANT: Howard, Bussey					
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery					
; FILE REFERENCE: 10102-005-999					
; CURRENT APPLICATION NUMBER: US/10/032,585					
; CURRENT FILING DATE: 2001-12-20					
; SOFTWARE: PatentIn version 3.1					
; SEQ ID NO 6493					
; LENGTH: 1239					
; TYPE: DNA					
; ORGANISM: Candida albicans					
US-10-032-585-6493					
Query Match					
Best Local Similarity 99.8%; Pred. No. 2,6e-227;					
Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
QY					
1 ATGAGTGAAGTCAGCAACCAATGCATATCTTTATATCTTCTTCATCA 60					
Db					
1 ATGAGTGAAGTCAGCAACCAATGCATATCTTTATATCTTCTTCATCA 60					
QY					
61 CGTCCCAAAAGTATTTTGACATATGGAAGGTGTGTAAGCCATATATGACCATCA 120					
Db					
61 CGTCCCAAAAGTATTTTGACATATGGAAGGTGTGTAAGCCATATATGACCATCA 120					
QY					
121 TTATTAGACACATTTAAGAACCCAGTATGATGACCGATTAATGTAAGTGACAGTGAC 180					

```
Db      121 TTAATGGAGCAACATTTAAGAACCCACAGTAATGACGACCGATATAATGTATACAGGAG 180
QY      181 GATTGTGATTAAGACATTTTTCAGAAAATCACTTTTGGAAAACATATTTGTATCACTTCC 240
Db      181 GATTGTGATTAAGACATTTTTCAGAAAATCACTTTTGGAAAACATATTTGTATCACTTCC 240
QY      241 GAAAAAAAACATTTCCATTTGTCAGTGTGTGTAAAGGGTTAAATCTCGACAAACATTTG 300
Db      241 GAAAAAAAACATTTCCATTTGTCAGTGTGTGTAAAGGGTTAAATCTCGACAAACATTTG 300
QY      301 AAAAGACATGAATCACTTACAAAGTCACTTTAAATGTATCATTTGAAAAATTTGTCAAGA 360
Db      301 AAAAGACATGAATCACTTACAAAGTCACTTTAAATGTATCATTTGAAAAATTTGTCAAGA 360
QY      361 GCATTTTAAATCATCAATCTTTAAGACATCATATTTATCTGTGATGAAAAAACAATTA 420
Db      361 GCATTTTAAATCATCAATCTTTAAGACATCATATTTATCTGTGATGAAAAAACAATTA 420
QY      421 ACSTGTAAACAATGTATAAAGTTTCACTCGACCTTCAAAATTAAGACAACTAAATTA 480
Db      421 ACSTGTAAACAATGTATAAAGTTTCACTCGACCTTCAAAATTAAGACAACTAAATTA 480
QY      481 AAAACATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGTGTTTAAAAATTTTC 540
Db      481 AAAACATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGTGTTTAAAAATTTTC 540
QY      541 CAAACTGTGCAATTTTCAATTTCACTATAAACAATGATCCAAACCTTAATGCTC 600
Db      541 CAAACTGTGCAATTTTCAATTTCACTATAAACAATGATCCAAACCTTAATGCTC 600
QY      601 AAATGTGTAAAGTGTGTGTGGAAAAAGTTTATCTTCACTATGTATTAAGTCATGAT 660
Db      601 AAATGTGTAAAGTGTGTGTGGAAAAAGTTTATCTTCACTATGTATTAAGTCATGAT 660
QY      661 GATTCTACATGATCAAAATATGACTGTGATTAATGTGATGTGGGAAATTTGCAAG 720
Db      661 GATTCTACATGATCAAAATATGACTGTGATTAATGTGATGTGGGAAATTTGCAAG 720
QY      721 AAAAATGAATTAAGTGAACATTAATATCTTCATGATGTGATTAATCTCCGTGATGATTA 780
Db      721 AAAAATGAATTAAGTGAACATTAATATCTTCATGATGTGATTAATCTCCGTGATGATTA 780
QY      781 TTAAGGAAACTGAAGTGAAGAAATTAAGAACCTTATTAAGATCAAGATGAAATTAAT 840
Db      781 TTAAGGAAACTGAAGTGAAGAAATTAAGAACCTTATTAAGATCAAGATGAAATTAAT 840
QY      841 AATTGTCATGATTAAGAACAGAGAAATTAAGAGTGAAGAGATGAAGAA 900
Db      841 AATTGTCATGATTAAGAACAGAGAAATTAAGAGTGAAGAGATGAAGAA 900
QY      901 GATATGCTATGATGAAGAAAGTGAATGTATGATCAAGTCAATGTCAAGTCAAAATCA 960
Db      901 GATATGCTATGATGAAGAAAGTGAATGTATGATCAAGTCAATGTCAAAATCA 960
QY      961 ATTAATCAATTTACTGCTTCTTGAAGGTTCAAGAGTGTCTAAACTTATTCGAAT 1020
Db      961 ATTAATCAATTTACTGCTTCTTGAAGGTTCAAGAGTGTCTAAACTTATTCGAAT 1020
QY      1021 AGTGGGAAGAGATCAATGTCTTAAGAAATTAATGTGATTAAGATGTTTTCTAGAGATAT 1080
Db      1021 AGTGGGAAGAGATCAATGTCTTAAGAAATTAATGTGATTAAGATGTTTTCTAGAGATAT 1080
QY      1081 GATTTCAGTGAACATTTGAAGTGCATGATGATTAATTTACAAAGATTTGATCTTATTA 1140
Db      1081 GATTTCAGTGAACATTTGAAGTGCATGATGATTAATTTACAAAGATTTGATCTTATTA 1140
QY      1141 AATAGATTAAGAAAAAGAGAACTCCAGAGGTGAACCATTTGGTTAAAAAGCCAGAGATG 1200
Db      1141 AATAGATTAAGAAAAAGAGAACTCCAGAGGTGAACCATTTGGTTAAAAAGCCAGAGATG 1200
QY      1201 GATTATTTGCCAAATGAACATCAAGTATTTCTCGATTA 1239
Db      1201 GATTATTTGCCAAATGAACATCAAGTATTTCTCGATTA 1239
```

```
Db      1201 GATTATTTGCCAAATGAACATCAAGTATTTCTCGATTA 1239

RESULT 2
US-10-094-749-726
; Sequence 726, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAMA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL, FULL-LENGTH cDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: Patencin Ver. 2.1
; SEQ ID NO 726
; LENGTH: 2132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-726

Query Match      8.0%; Score 99.4; DB 17; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4e-09;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY      7 GAAAGTGACGAACCAATTCATATCATCTTTAATATCTTCTTCTTTCATCAAGTCCC 66
Db      1384 GATGTGGCAAGCTTTTAAACCAAGTCTTCAATCTTACTACACATAGAATTCATACT 1443
QY      67 AAAAGTATATTGACATATGAGAGGTGTATAAGCCTATATGACCATCATTAATTA 126
Db      1444 GAGAGAAATCTTCAAAATGTAAAGATGTGGCAAGCTTTCTATCGATCTTCAAACTT 1503
QY      127 GAGCAACATTTTAAGAACCCACAGTATGATGACCGTATTAATGTACATGAGCAATGT 186
Db      1504 ACTGAACATTAAGAAATTTATATCTGAGAGAAACCTTACATGT-----GAAAGATGT 1557
QY      187 GATTAAGCATTTTTCAGAAAAATCACTTTGGAACACATATTTGTATCATATTCGAAAAA 246
Db      1558 GCGAAAGCCTTTTAAACCACTCTTCAACCTTGTACATTAAGTAAATTCATCTGAGAG 1617
QY      247 AAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATCTCGAACACTTGAAGAA 306
Db      1618 AAACCTTCAATGTGAAGAAATGTGTAAAGCCTTTAAACAGTCTTCAACCTTACTAGA 1677
QY      307 CATGAATCACCCATACAAAGTCATTTAATGTATCAATTTGAAATTTGTCAAGAGCATTT 366
Db      1678 CATTAAGAAATTCATATCTGAGAGAAACCTTCAATGTGAAGAAATTTGTGCAAGCTTT 1737
QY      367 TATTAACATCAATCTTTAA--GACATCATATATTTATCTGTTCATGAAGAAAAATTAACG 423
Db      1738 AACCACTCTCAAACTTACTGACATTAAGAAATTCATATCTGTGAGAAACCTTACAAA 1797
```

OY		424	TGTAACAATGTAAATAAGTTTCACCTCGACCTTCAAAAATTAGACAAACATTAATTAAAA	483
Db		1798	CCTAAGAATGTAAAGATTTTGGAACCTTCAAAGTTTTCTAAACATTAAGAAAT	1857
OY		484	CATCATGGTGCATCTCCTGCTTATCAATGATGATCATCTGGTGTGTTTAAATAATCCAA	543
Db		1858	TATGCTGGTGAAGAAATCTTAGAAATGTGAAGAAATGAACAAACCTTTAAAGTTGCCAC	1917
OY		544	ACTTGTCAGTAATTCATTAATTTCAATTAATAACAATGCAATCCAAT	590
Db		1918	ACTTGATGTGCATATAGATTAATTCATTAATAAAAAAAAAAACCCTATAGT	1964
<b>RESULT 3</b>				
		US-10-305-720-1054		
		; Sequence 1054, Application US/10305720		
		; Publication No. US20040010136A1		
		GENERAL INFORMATION:		
		; APPLICANT: Au-Young, Janice K.; Sellhammer, Jeffrey J.		
		; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expression		
		; FILE REFERENCE: PA-0002-1 CON		
		; CURRENT APPLICATION NUMBER: US/10/305,720		
		; CURRENT FILING DATE: 2002-11-26		
		; PRIOR APPLICATION NUMBER: 09/016,434		
		; PRIOR FILING DATE: 1998-01-30		
		; NUMBER OF SEQ ID NOS: 1490		
		; SOFTWARE: PERL Program		
		; SEQ ID NO 1054		
		; LENGTH: 2320		
		; TYPE: DNA		
		; ORGANISM: Homo sapiens		
		FEATURE:		
		; NAME/KEY: misc feature		
		; OTHER INFORMATION: GenBank ID No. US20040010136A1 g1017721		
		US-10-305-720-1054		
		Query Match	7.8%; Score 96.6; DB 17; Length 2320;	
		Best Local Similarity 51.3%; Pred. No. 1.4e-08;		
		Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;		
OY		69	AAAGTATATTGGCACATATGAAAGGTGTGATAAAGCCATATATGACCATCATATTAGA	128
Db		964	AGAGAACCCCTACAAATGTAAAGATGTGTAAAGCTTTAAACGATCTTCAACCCCTAC	1023
OY		129	GCAACATTTAAGAACCCACAGTAATGATCGACCGTAAATATGTAAGTGACATGTGA	188
Db		1024	TACCCATAGAAAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAGATGTGG	1077
OY		189	TAAAGCATTTTTCAGAAAAATCACATTTGGAAACACATATGTATCATATTCGAAAAAAA	248
Db		1078	CAAGCCTTTAGAGAGCTCTCAACCTTACTACACATTAAGTATTAATCTAGTGAAGAA	1137
OY		249	ACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTTCTCGACAACCTTGAAGACAC	308
Db		1138	ACCCCTACAAATGTAAAAAATGTGAAAAAGCCTTTTACCAAGCTGCACACCTTACCAACA	1197
OY		309	TGAATACCCCATTAAGATTTAAATGTACATTTGAAAAATGTGAAGAGACTTTTA	368
Db		1198	TGAGGTATTAATTAATCTGAGAGAAACCTTACAAATGTGAAGAAATGTGAAGAGCTTTAA	1257
OY		369	TAAACA--TCATCTTTAAGACATCATATTAATTAATCTGTTTCAATGAAAAAATTAACGTG	425
Db		1258	TCATTTTCACACCTTACTACATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1317
OY		426	TAAACATGTATTAAGTTTTCACCTGACCTTCAAAATTAAGACACATTAATTAATA	485
Db		1318	TAAAGAAATGTGTAAAGCTTTTAAACACTTTCACACCTTACTAAACATTAAGATTAATCA	1377
OY		486	TCATGTGTGATCTCTGCTTATCAATGTGATCATCTGGTGTGTTTAAATAATTTCCAAC	545
Db		1378	TACTGAGAGAAAGCTTACATTAATTAAGATGTGAAGAAAGCTTTTAAACCAATCTCATA	1437

Oy	546	TTGGTCAGTATTAACAATTCTCATATAAACAACCTGATCAACAACTTAAATGTCTCAATG	605
Dd	1438	ACTTAGACATGAATGAAAAATTTCACTCGAGAGAAACCCTA---TGAAATGAAAAATG	1494
Oy	606	TGCTAAAGCTTGTGTGGGAAAAAGGTTATATCTTCAATATGTAAATGCATGATGA	662
Dd	1495	TGGCAAAGCTTTTAAACAGTCCCTCAATCTTACTAGACATTAAGAAGTCATTACAGA	1551
 RESULT 4 US-10-029-386-20763/c ; Sequence 20763, Application US/10029386 ; Publication No. US20030194704A1 ; GENERAL INFORMATION: ; APPLICANT: Penn, Sharon G. ; APPLICANT: Rank, David R. ; APPLICANT: Hanzel, David K. ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO ; FILE REFERENCE: ABOVICA-X-2 ; CURRENT APPLICATION NUMBER: US/10/029,386 ; CURRENT FILING DATE: 2001-12-20 ; NUMBER OF SEQ ID NOS: 34288 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1 ; SEQ ID NO 20763 ; LENGTH: 994 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; OTHER INFORMATION: MAP TO AC010620.3 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4 ; OTHER INFORMATION: NT HIT: g13650821, EVALUO 0.00e+00 ; OTHER INFORMATION: EST HUMAN HIT: B1093435.1, EVALUO 0.00e+00 ; OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUO 0.00e+00 US-10-029-386-20763			
 Query Match      7.4%; Score 92; DB 16; Length 994; Best Local Similarity    52.0%; Pred. No. 7.7e-08; Matches    259; Conservative    0; Mismatches 220; Indels    9; Gaps    2;			
Oy	7	GAAAGTACGAACCAATGCATATCATCTTTAAATATCTTCTTCTTCAACAGTCCC	66
Dd	763	GAAATGTGGCAAAAGCCTTCTACCATCTTTCACACCTTCTACACATTAAGSTAAATTCATACT	704
Oy	67	AAAAGTATATTGGCATATATGAAGSGTGATGAAGCATATATGACCATCATTTATTA	126
Dd	703	GGAAGAAAGCCCTTCAAAATGTGAAGAAATGTGTAAGACTTTTAAACACCCTTCAGCCCTT	644
Oy	127	GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTAAATGTATACAGTGAAGCATGT	186
Dd	643	ACTACACATTAAGTTCATTTATGTTAAAGAAAAAACCTTACAATGT-----GAAGAAATGT	590
Oy	187	GATAAAGCATTTTTCAAAAAATCACATTTGGAAAACATATTTGATACATTTCCGAAAA	246
Dd	589	GACAAACCTTTTAAACCATTTCTCATACCTTACTTAAACATTAAGATTAATTCATTTCTGGAG	530
Oy	247	AAACCATTTCAATGTTGATGTGTGTGAAGGGGTAAATCTCGAACAACTTGAAAGA	306
Dd	529	AAATCTTCAAAATGTGAACATGTGGCAAGGCTTTAACTGTCTTCAACCCCTTACAAA	470
Oy	307	CATGAATCAACCCATCAAAAGTCATTTAAATGTATCATTTGAAAATTTGTCAAGAAAGATT	366
Dd	469	CATAAAGAATTCATATCGGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCCTTT	410
Oy	367	TATTAACAT---CAATCTTTAAGCATCATATATATCTGTTCATGAAAAAACATTACG	423
Dd	409	AATGTGTCTTCAACCTTACTACACATTAAGATGATTCATCTGTGAGAGAAACCTTACAA	350



```

; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
; LENGTH: 3078
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-622

```

```

Query Match      7.4%; Score 91.8; DB 17; Length 3078;
Best Local Similarity 52.1%; Pred. No. 1.3e-07;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

```

```

QY 11 GTGACGAAACCAATGATATCATCTTAATATCTTCTTCTGATCAGTCCGAA 70
DB 650 GTGCAATATATTTTGATGATCTTTCACATTAATCACTCACTAAGAAATTCATCTAG 709
QY 71 AGTATATTTGACATATGAAGGGTGTGATGAAGCTTAATTCAGCATCATTTAGAGC 130
DB 710 AGTATCTTCAATATGGAAGATGTGTAAAGCTTTAATCTGCTCTCAACCTTACTA 769
QY 131 AACATTTAGAACCCACAGTAATGATGACCGTAATAATGTACAGTGGACGATTGTGATA 190
DB 770 AACATAGATTAATTCATCTACTGAGAAACCCCTACAAATGT-----GAAGATGTGGCA 823
QY 191 AACATTTTTCAGAAATATCATTTTGGAAACACATATGTATTCATTCGGAAGAAAC 250
DB 824 AACCTTTTAAACCGGTCTCTCAATCTTCTTAACATTAATTAATTCATCTGAGAGAAC 883
QY 251 CATTCATTTGTTCAGTGTGTGAAGGGGTAAATTCCTGACAACTTGAAGAACATG 310
DB 884 CCAACAAATGTGAAGATGTGGCAAGCTTTAAACGGTCTTCAACCTTACTTAACATA 943
QY 311 AATACACCATACAAAGTCATTTAAATGATCATTTGAAATTTGTCAAGAAACATTTTATA 370
DB 944 AAGAAATTCATACAGAAAGAAACCCCTACAAATGTGAAGATGTGGCAAGGCTTTAAC 1003
QY 371 AACATCAATCTT---AAGCATCATATATATCTGTCTGAAAAACATTAAGCTGA 427
DB 1004 AGTCTCGATTTTAAATTAACATTAAGAAATTCATATGGAAGATTAACCTTACAAATGTG 1063
QY 428 AACATGATATTAAGTTTTCATCTGACCTTCAAAATTAGCAACAATTAATAAACATC 487
DB 1064 AAGAAATGTGGCAAGCTTTAGAGTATCTCAATTTCTTAATAAAACATTAAGATATCATTA 1123
QY 488 ATGTGATCTCC 500
DB 1124 CTGGGGAAGAAAC 1136

```

```

RESULT 8
US-10-094-749-692
; Sequence 692, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:

```

```

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORI
; APPLICANT: NAGAI, KEIICHI

```

```

; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKI, ICHIRO
; APPLICANT: SEKI, MACHIRO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 692
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-692

```

```

Query Match      7.3%; Score 90.4; DB 17; Length 2114;
Best Local Similarity 51.8%; Pred. No. 2.1e-07;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

```

```

QY 7 GAAAGTGACGAACCAATGATATCATCTTTAATATCTTCTTCTTATCATCGTCC 66
DB 1024 GAATGTGGCAAGCTTCTTCAATCTTCTTCAACCTTACTACATTAAGTAATTCATCT 1083
QY 67 AAAAGTATATTTGACATATGAAGGGTGTGATGAAGCTTAATGACATCATATTA 126
DB 1084 GAGAGAACCCCTTCAATATGTGAAGATGTGTAAAGCTTTTAAACCACTTCAAGCCTT 1143
QY 127 GAGCAATTTTAAAGAACCAAGTATGATGACCGTATTAATGTACAGTGACGATTGT 186
DB 1144 ACTACACATTAAGTCTTCTATCTGTTAAAGAAAAACCTTCAAAATGT-----GAAGATGT 1197
QY 187 GATTAAGCATTTTTCAGAAATTCATTTGAAACACATATTTGATCATCTCGAAAA 246
DB 1198 GACAAAGCTTTTAAACGATTTCTCATCTTACTTAACATTAAGTAATTCATCTCGAGAG 1257
QY 247 AAACATTCATTTGTTCAGTGTGTGAAGGGGTAAATTTCTGCAACAACCTTGAAGA 306
DB 1258 AATCTTCAAAATGTGAACAAATGTGGCAAGGCTTTAAGCTTCAACCTTACAAAA 1317
QY 307 CATGAATACCACTACAAAGTCATTTAATGTACATTTGAAATTTGTCAAGAAAGATT 366
DB 1318 CATGAAGAAATTCATCTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCCTTT 1377
QY 367 TATTAACAT---CAATCTTAAAGATCATATATATCTGTTCATGAAAAACATTAAG 423
DB 1378 AATGTCTTCAACCTTACTACATAAGATGATTCATATCGAGAGAAACCTTACAAA 1437
QY 424 TGTAAACATGTATTAATTAAGTTTTCATCTGACCTTCAAAATTAGCAACAATTAATTA 483
DB 1438 TGTGAAGATGTGGCAAGCCTTTAACAACCTTCAAACTTACTAATATTAAGATTAAT 1497
QY 484 CATCATGTGATCTCCT 501
DB 1498 CATCTGGAGAGAAACCT 1515

```

```

RESULT 9
US-10-094-749-499

```

```

; Sequence 499, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI

```

```
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YURI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKI, ICHIRO
/ APPLICANT: SEKI, NAOHITO
/ APPLICANT: YOSHIKAWA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOMYUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
/ FILE REFERENCE: 084335/0160
/ CURRENT APPLICATION NUMBER: US/10/094,749
/ PRIOR FILING DATE: 2002-03-12
/ PRIOR APPLICATION NUMBER: 60/350,435
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: JP 2001-328381
/ PRIOR FILING DATE: 2001-09-14
/ NUMBER OF SEQ ID NOS: 3381
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 499
/ LENGTH: 2230
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-094-749-499
```

```
Query Match 7.3%; Score 90.2; DB 17; Length 2230;
Best Local Similarity 50.4%; Pred. No. 2.4e-07;
Matches 332; Conservative 0; Mismatches 313; Indels 14; Gaps 4;
```

```
QY 7 GAAAGGAGAAACCAATCGATATCATCTTTAATCTCTTCTTCTCATACGTC 66
DB 656 GAATGAGCAATACACTTGGCATGCTTTCACCGCTTCACTCAATAAAAATTCATCT 715
QY 67 AAAAGTATATTTGACATATGAGGGGTGATTAAGCCATATGACCATATTA 126
DB 716 AAGAGAAATTTCTCAAAATGTGAAGGTGTGAAAACTTTTACTGCTCACAACCTT 775
QY 127 GAGCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGA 186
DB 776 TCTAACTTAAGAAATTCATCTGAGAAAAACCTTACAAATGTGAAGTA-----TGT 829
QY 187 GATTAAGCATTTTTCAGAAAAATCATCTTTGAAAAATGTATCATCTCCAAAA 246
DB 830 GGAAGACCTTTTCAACCAATCTCAATCTTAACTTAACATAGATTAATTCGTACTGAGAA 889
QY 247 AAACATTCATGTTCAAGTGTGTTAAGGGGTTAATTTCTGACAAACATTTGAAAGA 306
DB 890 AAACCTTAATTAATGTGACACGTGTGCAAAAGCCTTTAAACAGTCTTCAACCTTACTAGA 949
QY 307 CATGAATGACCCATCAAAAGTCATTTAATGTATGACATTTGAAATTTGCAAGAGCATTT 366
DB 950 CATTAATATTCATCTGAGAGAAACCTTCAAAATGTGACAAATGTGGCAAGCTTT 1009
QY 367 TATTAACATCATC---TTTAAGACATCATATATATCTGTTTCATGAAAAAAACATTAGC 423
DB 1010 AAGAGTCCCAACCTTCAACATGATTAATTAATCTGAGAGAAACCATTCACAA 1069
QY 424 TGTAAACATATTAATTAAGTTTCACTGACCTTCAAAATTTAGCAAAATTAATTAATA 483
DB 1070 TGTGAGAAATGTGCAAGCTTTTAACATTAATTAACCTTAAGCAATTAAGAAATTT 1129
QY 484 CATCATGTGTG-GATCTCTGCTTATCATATGTGATCATCTGTTGTTTAAATAATTTCA 542
DB 1130 TACACTAGAGAGAAACCTTACAAATGTGAAGATGTGGAAAGCCTTTTAAACAGTTTGA 1189
QY 543 AACTGTGTCAGATTACAAATTTCAATTAATAACAACTGATCCAAAACTTAAATGTCTTA 602
DB 1190 ACCCTTATTAACATTAAGATTAATTCATAGACAGAGAAACCC---CACAAATGTGAAGA 1245
```

```
QY 603 ATGTGTAAAGGTTGTGGAAAAAGGTTTATCTTCAATATGTAAGTCATGATG 661
DB 1246 ATGTGGCAGAGCTTTTAACCAAGTCCGCAAGCTCATGAATTAATTAATCACTG 1304
```

## RESULT 10

```
US-10-029-386-13298
/ Sequence 13298, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
/ FILE REFERENCE: AEWICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ PRIOR FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 13298
/ LENGTH: 590
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO CHR19.3
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
/ OTHER INFORMATION: NT HIT: 91475315, EVALUR 0.00e+00
/ OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALUR 1.00e-119
/ OTHER INFORMATION: EST_HUMAN HIT: BG619358.1, EVALUR 0.00e+00
US-10-029-386-13298
```

```
Query Match 7.2%; Score 89.8; DB 16; Length 590;
Best Local Similarity 53.8%; Pred. No. 1.7e-07;
Matches 234; Conservative 0; Mismatches 192; Indels 9; Gaps 2;
```

```
QY 62 GTCCAAAAAGTATTTTGCATATGAGGGGTGATTAAGCCATATGACCATCAT 121
DB 126 GTGCTTAAGAGAAATTTCTACAGTGTGAAGATGTGGCAACATTTCTATGCTCTCA 185
QY 122 TATTAGACCAATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAGC 181
DB 186 CCTTAAGTACATTAAGAGATACACACTGGAGAAACCTTCAAAATGT-----GAAG 239
QY 182 ATGTGATTAAGCATTTTTCAGAAAAATCATCTTTGAAAAACATATTTGATCATCTCG 241
DB 240 AATGTGGCAAGCTTTTGTAGCCATTTCTCAACCTTGTAAACATTAAGAAATTCATCTG 299
QY 242 AAAAAAACCATTCATGTTCAAGTGTGTTAAGGGGTTAATTTCTGACAAACATTTGA 301
DB 300 GAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTAGCCGTTCTTCAACCTTG 359
QY 302 AAAGCATGAATATCCCATCAAAAGTCATTTAATGTATGACATTTGAAATTTGTCAAGAG 361
DB 360 CTAACATTAAGATTAATCTACTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAG 419
QY 362 CATTTTATTAACATCATATCTTTAAGACATCATATATATCTGTTTCATGA---AAAAAT 418
DB 420 CTTTATGAATTCCTCAACCTTGTCAATTAATTAATCAATCAATCAATCAAGAAACCT 479
QY 419 TAACTGTAAACAAATTAATTAAGTTTCACTGACCTTCAAAATTTAGCAAACTTAAT 478
DB 480 ACAATATTAAGAAATGTGAACAAAGCTTTTAAGCGACTCTCAACCTTACTTAACATTA 539
QY 479 TAAACATCATGTGTG 493
DB 540 TATATCATGCTGGAG 554
```

```
RESULT 11
US-10-108-260A-29
```



```

; Sequence 29, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 2221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-29

Query Match      7.2%; Score 89.2; DB 17; Length 2221;
Best Local Similarity 53.4%; Pred. No. 3.7e-07;
Matches 237; Conservative 0; Mismatches 198; Indels 9; Gaps 2;

QY 61 CGTCCCAAAAGATATTTGACATATGAAAGGTGTGATTAAGCCTATATGACCATCA 120
DB 654 CATACCGAGAGAAACCTTTCAGATGTAAAGAAATGGCAAGCCTTTATAGAACTCA 713
QY 121 TTATTAAGGCAATTTAAGACCCAGTAATGATCGACCGTATTAATGTACAGTGGAC 180
DB 714 TACCTTACTCAGATCAGAGAAATGCTATGAGAGAAACCTTTCAAGTGT-----AG 767
QY 181 GATTGTGATTAAGCATTTTTCAGAAATACATTTGGAAACACATATTTGTACATTC 240
DB 768 GAATGGGCAAACTTTTAAGAGGCTTCAACCTTACATCAAGAGATCAAGAAATCCAT 827
QY 241 GAAAAAAACCATTTCCATTTGTCACTGTGTGTAAAGGGTTAATTTCTGCAACACTTG 300
DB 828 GGAGAGAAACCTTCAATGTAAAGATGTGGCAAGCCCTTTAAACGGGGCTCACACTT 887
QY 301 AAAAGCATGAATTCACCCATACAAAGTCAATTAATGTCAATTTGAAATTTGCAAGAA 360
DB 888 ACTCGACATCAAGAAATCTACTGAGAGAGAGTCTTTCAAAGTGAAGATGTGGCAAA 947
QY 361 GCATTT---TATTAACATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAACA 417
DB 948 GCTTTATCTGGGGCTCACACCTTACATCAATGAGAGTCCATCTGAGAGAAATTC 1007
QY 418 TTAACTGTAAACATGTAAATTAAGTTTCACTGACCTTCAAAATTAAGCAACATAAA 477
DB 1008 TTCAAAATGTAAAGATGTGGCAAGCTTTTACAGAGAGTCAACCTTACTCAACATCAG 1067
QY 478 TTAAACATCATGTGTGATCTCCT 501
DB 1068 AGAATTCATCTGAGAGAAACCT 1091

RESULT 12
US-10-172-118-910
; Sequence 910, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; TYPE: DNA

```

```

; SEQ ID NO 910
; LENGTH: 3839
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003430
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-910

Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAATATATTTGACATATGAAGGGTGTATTAAGCCTTAATTCGACATCATTTATTAGA 128
DB 964 AGGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTACCATCTTCAACCTTCC 1023
QY 129 GCACATTTAAGAACCCACAGTAAATGATGACCGTAAATGTATACAGTACGATTTGA 188
DB 1024 TAAACATAAGAAATTCATATCTGAGAGAAACCTTCAAAATGT-----GAAGATGTGG 1077
QY 189 TAAAGATTTTTCAGAAATTCATTTGAAACACATATTTGTATCATTTCCGAAAAAA 248
DB 1078 CAAAGCTTTTACCATTTCTTCAAGCCCTTGTAAACATTAAGAAATTCATATCTGAGAGAA 1137
QY 249 ACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCGACACACTTGAAAGACA 308
DB 1138 ACCCTACAAATGTAAAGATGTGGCAAGCTTTTACAAATTCCTCAACCTTGTATCA 1197
QY 309 TGAATACCCATACAAAGTCAATTAATGTACATTTGAAATTTGTCAAGAAAGATTTTA 368
DB 1198 TAAGTAATCATATCAGAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTA 1257
QY 369 TAAACATCAATC---TTTAAGACATCATATTTATTTGTTCATGAAAAAACAATTACG 425
DB 1258 GGCACCTCAACCTTACTTAAACATTAATTAATCATGTGTGAGAGAAACCTTCAAAAG 1317
QY 426 TAAACATTAATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTAACA 485
DB 1318 TGAAGATGTGGCAAGCTTTTATGATCTTCAAAATCTTACTATACATTAAGTTTATCA 1377
QY 486 TCATGTGTGATCTCCT 501
DB 1378 TACTGAGAGAAACCT 1393

RESULT 13
US-10-342-887-910
; Sequence 910, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
; LENGTH: 3839
; TYPE: DNA

```

```
/ ORGANISM: Homo sapiens
US-10-342-887-910

Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAATATATTTGACATATGAGGGGTGTATTAAGGCTATTAATGACCATCATTTATTA 128
DB AGGAAACCCCTCAAAATGTGAAGAAATGTGGCAAAAGCTTTAGCCATTCTTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCAGATATATGATGACCGGTATTAATGACGTGACATTTGCA 188
DB 1024 TAAACATAAGAGATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAATGTG 1077
QY 189 TAAAGCATTTTTCAGAAAATCACATTTGGAAAACATATTGTATCATTTCCGAAAAAAA 248
DB 1078 CAAAGCTTTTACCATCTTCTGAGCCCTTGCTAAACATTAAGAAATTCATCTGAGAGAA 1137
QY 249 ACCATTCATTTGTCAGTGTGTGTAAAGGGGTTAAATTCGACAAACATTTGAAAAGACA 308
DB 1138 ACCCTTCAAAATGTGAAGATGTGGCAAAAGCTTTTACAAATTCCTCAACCTTGCTAATCA 1197
QY 309 TGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAACATTTTA 368
DB 1198 TAAAGTAACTCAATCTGAGAGAAACCTTCAAAATGTGAAGATGTGACAAACCTTTTAA 1257
QY 369 TAAACATCATC---TTAAGACATCATATATATCTGTTCATGAAAAAAACATTAACGTG 425
DB 1258 GGAATCTCAACCTTACCTTAACATTAATTAATCATGTGTGAGAGAAACCTTACAAAG 1317
QY 426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTAAGCAACATTAATTAATAACA 485
DB 1318 TGAAGATGTGGCAAGCTTTTATGATCTTCAATCTTACTATCATTAAGTTTATCA 1377
QY 486 TGAATGTGATCTCCT 501
DB 1378 TACTGAGAGAAACCT 1393

RESULT 14
US-10-098-841-48
/ Sequence 48, Application US/10098841
/ Publication No. US20020197679A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenshua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Xu, Chongjun
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yungqing
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Ren, Feiyun
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Zhang, Jie
/ APPLICANT: Qian, Xiaohong B.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: No. US20020197679A1e1 Nucleic Acids and
/ FILE OF INVENTION: Polypeptides
/ FILE REFERENCE: 784CIP2
/ CURRENT APPLICATION NUMBER: US/10/098,841
/ CURRENT FILING DATE: 2002-03-13
/ PRIOR APPLICATION NUMBER: 09/598,042
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 331
```

```
/ SOFTWARE: pt_FL_genes Version 1.0
/ SEQ ID NO 48
/ LENGTH: 3502
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (154)..(2583)
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(3502)
/ OTHER INFORMATION: n = a,t,c or g
US-10-098-841-48

Query Match      7.2%; Score 89; DB 13; Length 3502;
Best Local Similarity 51.7%; Pred. No. 4.8e-07;
Matches 256; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 79 TGCACATATGAAGGGTGTGATTAAGCCATATATGACCATCTATTATTAGAGCAATTTA 138
DB 2101 TACAAATGTGAAGATGTGGCAAAAGCTTTAAGTGTCTCAACCTTATCAATTAAG 2160
QY 139 AGAACCCACAGTATGATGACCCGTATTAATGTACAGTGAAGATTTGATTAAGCATTT 198
DB 2161 ATAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAATGTGCAAGCTTT 2214
QY 199 TTCAGAAAATCACATTTGAAAACACATATTGTATCATTCGAAAAAAACCATTCAT 258
DB 2215 AAAGTGTCTCAACCTTTTACACATTAAGATTAATTTACTGAGAGAAACCTTACAAA 2274
QY 259 TGTTCAGTGTGTGTAAAGGGGTTAATCTCGACAAACCTTGAAGAAAGCATGAAATCAC 318
DB 2275 TGTGAAAAATGTGGAAAGCTTTTAAACGACCTTCAACCTTATTTGAACATTAAGAAATTT 2334
QY 319 CATACAAAGTCAATTAATGTACATTTGAAAATTTGTCAAGAACATTT---TATAACAT 375
DB 2335 CATCTGAGAGAAACCTTCAAAATGTGAAGATGTGCAAAAGCATTTTACTATTTCTCA 2394
QY 376 CAATCTTAAAGCATCATATATTAATCTGTTCATGAAAAAAACATTAACGTGTAACATGT 435
DB 2395 CACCTTAATTAACATTAAGATTAATCTTAAGAGCAACCTTCAAAATGTGAAGATGT 2454
QY 436 AATAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTAATAATCATATGTGCA 495
DB 2455 GGCAAAGCTTTCAACCAATTAATCAACCTTCTACACATTAACAAATTCATCTGAGAG 2514
QY 496 TCTCTGTCTTATCATATGATCATCTGATTTTAAAAATTCAAACTGTGTCAGTA 555
DB 2515 AAATCTTCAAAACCTGAAAGATGTGACATGATTTTGAACACACCTCAACCTTTCAAC 2574
QY 556 TTACATTTTCATATA 570
DB 2575 ATTAATTAATTAATA 2589

RESULT 15
US-10-029-386-25074
/ Sequence 25074, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharron G.
/ APPLICANT: Rank, David K.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
/ FILE OF INVENTION: EXPRESSION ANALYSIS TWO
/ FILE REFERENCE: ABOMICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 25074
/ LENGTH: 1466
/ TYPE: DNA
/ ORGANISM: Homo sapiens
```

```

; FEATURE:
; OTHER INFORMATION: MAP TO AC011467.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
; OTHER INFORMATION: NT HIT: g1i5309152, EVALU8 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P35789, EVALU8 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AUI43734.1, EVALU8 0.00e+00
US-10-029-386-25074

```

```

Query Match      7.2%; Score 88.6; DB 16; Length 1466;
Best Local Similarity 53.3%; Pred. No. 4.1e-07;
Matches 237; Conservative 0; Mismatches 199; Indels 9; Gaps 2;

```

```

QY 60 ACCTCCCAAAAGTATATTTGCAATATGAGGGGTGATTAAGCCTATATGACCATC 119
Db 272 ACATACCTGGAAGAAATCTTCAATGTAAAGATGAAAGTCATTTGCAATGCTTC 331
QY 120 ATTATTAGACAATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGA 179
Db 332 ACCTTAGCTCAACATTAAGAAATTCATAGTGAAGAAACCTCAAAATGTAAA----- 386
QY 180 CGATTGATTAAGCATTTTTCAGAAATGACATTTGGAACAATATTTGATCATTC 239
Db 387 -GATGTGGGAAGCCTATATATGAGACCTCAACCTTCTACATTAAGAAATTCATAC 445
QY 240 CGAAAAAAACCATTCATTTGTCAGTGTGTAAAGGGTTAATTTCTGACACACTT 299
Db 446 TGGAAAGAAACCTTACAAATGCGAAGAGTGTGAAAAAGCCTTTAACGGGCTCAGACCT 505
QY 300 GAAAAGACATGAATCACCCCATACAAAGTCATTTAAATGTACATTTGAAAAATTGTCAGA 359
Db 506 TACTACACATTAAGTATATTCATCTAGTGAAGAAACCTCAAAATGTGAGAGTGTGCAA 565
QY 360 AGCATTTTAAACAT---CAATCTTAAGACATCAATATATCTGTTGATGAAAAAC 416
Db 566 AGCTTTTAACCAATCTGCAAAACCTTACTACACATTAAGAAATTCATACCTGAGAGAAACC 625
QY 417 ATTACGTGAACAAATGTAAATTAAGTTTCACTGCACTTCAAAAATTAGCACACATA 476
Db 626 CTACAAATGTGAAGATGTGGCAGAGCTTTTAGCCAGTCTCAACCTTTACTGCACTAA 685
QY 477 ATTAAAAATCATGTGTGATCTCTT 501
Db 686 GATTAATTCATGTGAGAGAAACCT 710

```

Search completed: February 9, 2005, 11:50:32  
Job time : 2162.11 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:45:39 / Search time 3815.62 Seconds  
(without alignments)  
12360.150 Million cell updates/sec

Title: US-09-831-804-2  
Perfect score: 1239  
Sequence: 1 atgagtgtaagagcagcgaac.....catcagtgattcttcgataa 1239

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	182.2	14.7	628	BZ298000	CG3957.f1
C 2	167.6	13.5	959	CNS060MP	AL405671 TT end of
3	103.8	8.4	801	CN759437	ID0AA25C
4	103.8	8.4	898	CN761003	ID0AA2AE
5	98.2	7.9	1609	3 AK032220	Mus muscu
6	97.8	7.9	2647	BC047646	Homo sapi
7	97.6	7.9	745	1 AU123448	AU123448
8	93.4	7.5	727	7 CO434402	UT-M-HXO-
9	92.4	7.5	631	7 CN352904	CN352904
10	92.4	7.5	2672	3 BC036394	Homo sapi
11	92	7.4	591	8 AQ005136	CIT-HSP-2
12	92	7.4	687	7 CN362485	170004241
13	92	7.4	752	7 CN417582	CN417582
14	91	7.3	593	8 B99387	CIT-HSP-228
15	91	7.3	820	5 BX414627	EX414627
16	90.8	7.3	666	7 CN297781	170005321
17	90.4	7.3	2694	3 BC025227	Homo sapi
18	90.4	7.3	2698	3 BC037782	Homo sapi
19	90.2	7.3	2253	3 BC028252	Mus muscu
20	89.8	7.2	550	5 BP292533	BP292533
21	89.8	7.2	582	5 BP234906	BP234906
22	89.2	7.2	604	2 BE161630	MR3-HT044
23	89.2	7.2	638	7 CN341035	CN341035
24	89.2	7.2	673	7 CN353930	170006001

25	89.2	7.2	3609	3 BC032590	BC032590 Homo sapi
26	89	7.2	2856	3 HSM803730	AL882422 Homo sapi
27	88.6	7.2	535	1 AL705393	DEF2686M
28	88.6	7.2	612	5 BP238960	BP238960
29	88.6	7.2	2174	3 AK033001	Mus muscu
30	88.6	7.2	2330	3 BC037426	BC037426 Homo sapi
31	88.6	7.2	2622	3 AK033958	Mus muscu
32	88.4	7.1	998	5 BX456765	BX456765
33	88	7.1	895	5 BU195416	AGENCOURT
34	87.8	7.1	608	7 CN266758	170004554
35	87.8	7.1	4251	3 CR749856	Homo sapi
36	87.6	7.1	744	5 BX437291	BX437291
37	87.2	7.0	579	8 AQ347265	AQ347265
C 38	87.2	7.0	687	8 AQ089266	RPC111-15
39	87.2	7.0	907	5 BQ423752	AGENCOURT
40	87	7.0	660	6 CD770140	CD770140
41	87	7.0	842	5 BU507408	AGENCOURT
C 42	86.8	7.0	521	8 AQ475626	CITB1-E1-
43	86.8	7.0	571	7 CN280110	170006001
44	86.8	7.0	705	8 AQ194282	RPC111-61
C 45	86.8	7.0	860	8 AQ749175	HS_5575_A

ALIGNMENTS

RESULT 1  
LOCUS BZ298000/c 628 bp DNA linear GSS 31-OCT-2002  
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata  
ACCESSION BZ298000  
VERSION BZ298000.1 GI:24440936  
KEYWORDS  
SOURCE GSS.  
ORGANISM Candida glabrata  
Candida glabrata  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
REFERENCE  
1 (bases 1 to 628)  
Wong,S., Faras,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.  
Evidence from comparative genomics for a complete sexual cycle in  
the 'asexual' pathogenic yeast Candida glabrata  
JOURNAL Genome Biol. 4 (2), R10 (2003)  
MEDLINE 22508158  
PUBMED 12620120  
COMMENT  
Contact: Wong S  
Department of Genetics, Smurfit Institute  
Trinity College Dublin  
Dublin 2, Ireland  
Tel: 353 1 6082319  
Fax: 353 1 6798558  
Email: swong@tcd.ie  
Classes: plasmid ends.  
Location/Qualifiers  
1..628  
/organism="Candida glabrata"  
/mol\_type="genomic DNA"  
/strain="CBS 138"  
/db\_xref="taxon:5478"  
/clone="CG3957"  
/clone\_lib="Candida glabrata Random Genomic Library"  
ORIGIN  
Query Match 14.7%; Score 182.2; DB 8; Length 628;  
Best Local Similarity 59.5%; Pred.No. 7.5e-26;  
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;  
QY 84 ATATGAGGCTGTGTAAGCCTATPATGACATCATTTATGACACATTTAAGAC 143  
DB 625 ATATGAGGCTGTGTAAGCCTATPATGACATCATTTATGACACATTTAAGAC 566  
QY 144 CCACAGTATGATCGACCGCTATTAATGTACAGTGTGATGATTAAGCATTTTCAG 203

Db 565 ---AGTCATCTTGGAGGAAACCTTGAAATGTAATCAATGTAAGTTCAATTACTAA 509  
 Qy 204 AAAATCACATTTGGAAACACATATGTAATCAGATTCGAAAAAACATTCATCTGTC 263  
 Db 508 AAAAGATCCACTAGAGAGACCTTATACACACAGATGAAAGCCGTTTATTTGTC 449  
 Qy 264 AGTGTGTGTAAAGGGGTTAATCTTCGACACACTTGAAGAATGATGAATCAACCATC 323  
 Db 448 ATTGTGTGGAAAGGGGCTTAATTAATGAGCAACAACTGAAACGATGAGTAATCAGAC 389  
 Qy 324 AAAGTCATTTAAATGATGATTTGAAATTTGCAAGAGATTTTAAATCAATCAATCTT 383  
 Db 388 CAATTCCTTCAATTTGTAATGAAATGAAAGTTGTAATGAGTTCTTCAAGAGACCTTCAAT 329  
 Qy 384 AAGACATCATATATTAATCTGTCATGAAAAAACATTAAGCTGTAACATGTAATAGT 443  
 Db 328 AAGGGCTCATATTTTATGAGATTCATTTACAAAGTCTAAATGTCAGATGACAAAG 269  
 Qy 444 TTTCACCTGACCTTCAAAATTAAGACACATTAATTAATAATCAATCAT-----GTTGATC 497  
 Db 268 TTTTCAAAAGCCCTTACAGACTCAAAATCAGATAGCAAAACACATTAATCCAGATGTTGT 209  
 Qy 498 TCCTGCTTATCATGATGATCCGTTGTTTAAATTTCCAAATCTGTCAGATTT 557  
 Db 208 CAATGCTTATCAATGATGATCTTCAATGTTGTTTCAAAACATGATGTCGCTT 149  
 Qy 558 ACAATTTCAATTAACAACTGATCCAACTCAAACTTAATGCTTAATGTAAGTAAAGTTG 617  
 Db 148 AAGATTACATGTTAAATGATCATCCGAAAGTTAAATGCCCTATATGACGTAACCTTG 89  
 Qy 618 TGTGGGAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662  
 Db 88 TGTGGGGAAGAGCGTTTAAATATGACATGAAATCCACGATGA 44  
 RESULT 2  
 CNS06MMP/c 959 bp DNA linear GSS 17-JUN-2001  
 LOCUS T7 end of clone AU00A005F10 of library AU00A from strain CBS 3082  
 DEFINITION of Saccharomyces kluyveri, genomic survey sequence.  
 ACCESSION AL405671  
 VERSION AL405671.1 GI:12168715  
 KEYWORDS GSS.  
 SOURCE Saccharomyces kluyveri  
 ORGANISM Saccharomyces kluyveri  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 REFERENCE 1 (bases 1 to 959)  
 AUTHORS Soucie,J., U., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malbertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Porter,S.,  
 Saurin,W., Tekala,F., Toffano-Nicche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 TITLE 2 (bases 1 to 959)  
 AUTHORS Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,  
 Gaillardin,C. and Casaregola,S.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 9.  
 JOURNAL FEBS Lett. 487 (1), 56-60 (2000)  
 MEDLINE 20584719  
 PUBMED 11152884  
 TITLE 3 (bases 1 to 959)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,  
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqef@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbicola*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.  
 FEATURES  
 source  
 location/Qualifiers  
 1..959  
 /organism="Saccharomyces kluyveri"  
 /mol\_type="Genomic DNA"  
 /strain="CBS 3082"  
 /db\_xref="taxon:4934"  
 /clone="AU00A005F10"  
 /clone\_11b="AU00A"  
 /note="end : 17"  
 /note="complement (<71..>958)  
 /note="similar to Saccharomyces cerevisiae ORF YPR186c [ PFI ; TFIIA (transcription initiation factor) ]"  
 /evidence=not\_experimental  
 misc\_feature  
 complement (<71..>958)  
 /note="similar to Saccharomyces cerevisiae ORF YPR186c [ PFI ; TFIIA (transcription initiation factor) ]"  
 /evidence=not\_experimental  
 ORIGIN  
 Query Match 13.5%; Score 167.6; DB 9; Length 959;  
 Best Local Similarity 57.2%; Pred. No. 6.1e-23;  
 Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;  
 Qy 211 CATTGGAAACATATTTGTAATCATATTCGAAACAAAAAACATTCATTTGTCAGTGT 270  
 Db 958 CACTTGGAAACATATTTGTAATCATATTCGAAACAAAAAACATTCATTTGTCAGTGT 899  
 Qy 271 GGTAAAGGGGTTAATTTCCGACACTTGAAGAGACATGAATACCATCAACAAAGCA 330  
 Db 898 GGTAAAGGGGTTAATTTCCGACACTTGAAGAGACATGAATACCATCAACAAAGCA 839  
 Qy 331 TTTAATGTATGATTTGAAATTTGCAAGAGACATTTTAAACATCAATCTTTAAGACAT 390  
 Db 898 TTTAATGTATGATTTGAAATTTGCAAGAGACATTTTAAACATCAATCTTTAAGACAT 779  
 Qy 391 CATATATTTATCTGTCATGAAACAACTTAACGTTAATCAATGATTAAGTTTCACT 450  
 Db 778 CATACACTTTCGTAACCTTCAGAAATTTGACCTGTGACATTTGGTAAATATTTCCA 719  
 Qy 451 CGACCTTGAATTAAGCAACATTAATTAACATCATGAGATCTCTGCT----- 504  
 Db 718 AGCCATATTCGTCGCAAAACATCTTGAAGACATCAATTTGATTTGAAACAA 659  
 Qy 505 TATCATGTGATCATCTCTGTTGTTTAAATTTCCAACTTGTGATTAACAATTT 564  
 Db 658 TATCATGTGATCATCTCTGTTGTTTAAATTTCCAACTTGTGATTAACAATTT 599  
 Qy 565 CATATTAACAACTGATCCAACTTAATGTCCTTAATGTTGTAAGGTTGTTGG 624  
 Db 598 CATATTAACAACTGATCCAACTTAATGTCCTTAATGTTGTAAGGTTGTTGG 539  
 Qy 625 AAAAAAGGTTTATCTTCATATGTTAAGTATGATGTTTACACATGATGAATATAG 684  
 Db 538 GAGTCTGCTTACAGAGTACATGATGATGATGATGATGATGATGATGATGATGATGAT 479  
 Qy 685 ACTTGTGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744  
 Db 478 AAATGTACATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419  
 Qy 745 AATATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 770  
 Db 418 ATGGAACATCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 393  
 RESULT 3  
 CN759437  
 LOCUS CN759437 801 bp mRNA linear EST 20-MAY-2004

DEFINITION ID0AAA25CC01RML ApMS Acyrthosiphon pisum cDNA clone ID0AAA25CC01 5', mRNA sequence.

ACCESSION CN759437

VERSION CN759437.1 GI:47533360

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.  
1 (bases 1 to 801)  
Hunter, W., Martinez-Torres, D., Rabhe, Y., Sabater-Munoz, B., Stern, D., Tagu, D. and Wincker, P.  
An expressed sequence tags database for the pea aphid Acyrthosiphon pisum  
Unpublished (2004)  
Contact: D. Tagu

JOURNAL INRA Rennes

COMMENT UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France  
Tel: +33.2.23.48.51.65  
Fax: +33.2.23.48.51.50  
Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope  
PCR Primers  
FORWARD: CAGGAACAGCTATGAC  
Plate: 25 row: C column: 1.  
FEATURES  
source  
1..801  
/organism="Acyrthosiphon pisum"  
/mol\_type="mRNA"  
/cultivar="developmentstage"  
/db\_xref="taxon:7029"  
/clone="ID0AAA25CC01"  
/issue\_type="whole insect"  
/dev\_stage="nymphs and adults (parthenogenetic females)"  
/lab\_host="XLI-Blue"  
/note="Vector: PBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ; Harvesting date: 01/06/1999 ; Stress date: no stress ; Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 801;  
Best Local Similarity 51.5%; Pred. No. 2,7e-10;  
Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

82 ACATATGAAGGGGTGTATAAGCTATATATCGACCATCATTTATTAAGCAATTATAGA 141  
Db AATATGATTTCTGTGTAAAGAGTTATCATATATACAGTTTAAACAGCATATCAAG 67

142 ACCACAGTATATGATGACCGGTATTAATGTACAGTGACGATTTGTATAAGCATTTTC 201  
Db ACACATATCCGGTGAAGACCGGTATGATGT-----GATACGTGTGATCAAGCGTTTCT 121

202 AGAAATACATTTGTAAGACACATATTTGATCATTTCCGAAAAAAAACCATTCATTTGT 261  
Db TTGAATATCATTTTAATATAGCCATTAATATGACATATCCGGCGAAAGCCCTTAATATGT 181

262 TCACTGTGTGTAAAGGGTTAATTCTCGACACACTTTGAAAGACATGAAATCACCAT 321  
Db GATTAACGTATCAAGGGTTTCTTAAGAAATCAAAATTTAACAAGGCATTAATAGACAT 241

322 ACAAAAGCATTTAATGTATCAATTGAAATTTGTCAAGACATTTTATAACATCAATCT 381

Db 242 ACCGGCAAAAGCCGTATTAATGTGAAAAGTGAACCAAGCGTTTCTCAGAAATCAGAT 301

Qy 382 TTAAGACATCATTTATATCTGTTTCAT---GAAAAACATTAACGTATAACATGTAAT 438  
Db 302 TTAATAGTCATATTAAGACACATACCGCGAAAGCCGATTAATATGTATACCTGAT 361

Qy 439 AAAGTTTACATCGACCTTCAAAATTAAGCACAACATTAATTAAGACATCATGCTGATCT 498  
Db 362 CAAGCGTTTTCATAAAATGATGATTTAATAGCATTAATTAAGACATACCGGTGAAGAG 421

Qy 499 CTTGCTTATCATATGATCATCTCTGTTGTTTAAAAATTTCCAAACTGGTCATATTA 558  
Db 422 C---CGTTAAATAGTCTAATCTGTATCAAGCGTTTATCGGAAAGCCATTTAAAAAGC 478

Qy 559 CAATTCATATTAACACATCGATCCAAACCTTAATATGCTTAATGTGTAAGGTGT 618  
Db 479 CACTCAAGACACATACAGGCCAAAAGCGTATTAATATGATACATGTAACCAAGCGTTT 538

Qy 619 GTTGGGAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTACATGATCAAA 678  
Db 539 TCTCAGAAATCAAAATTTAACAAGCCATACAGACACATGCCGTGAAGAGCTTTTAA 598

Qy 679 ATATGACTTGTGATTA 695  
Db 599 TGGATTAAGTGTATTA 615

RESULT 4  
CN761003 898 bp mRNA linear EST 20-MAY-2004  
LOCUS ID0AAA2AE05RML ApMS Acyrthosiphon pisum cDNA clone ID0AAA2AE05 5', mRNA sequence.

DEFINITION CN761003

ACCESSION CN761003

VERSION CN761003.1 GI:47534926

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.  
1 (bases 1 to 898)  
Hunter, W., Martinez-Torres, D., Rabhe, Y., Sabater-Munoz, B., Stern, D., Tagu, D. and Wincker, P.  
An expressed sequence tags database for the pea aphid Acyrthosiphon pisum  
Unpublished (2004)  
Contact: D. Tagu

JOURNAL INRA Rennes

COMMENT UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France  
Tel: +33.2.23.48.51.65  
Fax: +33.2.23.48.51.50  
Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope  
PCR Primers  
FORWARD: CAGGAACAGCTATGAC  
Plate: 2 row: E column: 5.  
FEATURES  
source  
1..898  
/organism="Acyrthosiphon pisum"  
/mol\_type="mRNA"  
/cultivar="developmentstage"  
/db\_xref="taxon:7029"  
/clone="ID0AAA2AE05"  
/issue\_type="whole insect"  
/dev\_stage="nymphs and adults (parthenogenetic females)"  
/lab\_host="XLI-Blue"  
/note="Vector: PBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ;



Harvesting date: 01/06/1999 ; Stress date: no stress ;  
Description: Aphids inoculated on one-week old *Vicia faba*  
under non-sterile conditions. All parthenogenetic stages  
and both winged and wingless adults were collected for  
library construction ; experimental condition: long  
photoperiod (16-hr light/8-hr dark at 18 °C)

## ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 898;  
Best Local Similarity 51.5%; Pred. No. 2.7e-10;  
Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

QY 82 ACATATGAAGGGTGTATTAAGCCCTTAATTCAGCATCATTTATAGACAACTTAAGA 141  
DB 7 AAATGATGTTTCTGTATTAAGAGTTTACTATATATCAAGTTTAAACAAGCAACAAG 66  
QY 142 ACCCAAGATATGATGACCGTATTAATGATGACGTGACGATTTGATTAAGATTTTC 201  
DB 67 ACACATACCGGTGAAAAAGCCGTATGATGT-----GATTAAGTATCAAGCGTTTCT 120  
QY 202 AGAAATACACATTGGGAAACACATATTGTATCACATTCGAAAAAAAACCATTCATGT 261  
DB 121 TTGAATACATTTATATAGCCATAAATGACACATACCGGGGAAAAGCTTATAATGT 180  
QY 262 TCAGTGTGTGTAAGGGGTTAATTCGACAACTTGAAGACATGAATACCCAT 321  
DB 181 GATTAAGTATGATCAACGTTTCTTAAGAAATCAAAATTTAAACAAGCATATTAAGACACAT 240  
QY 322 ACAAAATCATTAATGATCATTTGAATGTCAGAGACATTTATTAACATCATCT 381  
DB 241 ACCGGCGAAAAGCCGTATTAATGTGAAAAGTGAACCAAGCGTTTCTCAAAATCATGAT 300  
QY 382 TTAAGACATCATATATTTATCTGTTCAT---GAAAAACATTAACGTGTAACATGTAAT 438  
DB 301 TTAATACGTATATTAAGACACATACCGCGGAAAACCCGTATTAATGTATATCTGAT 360  
QY 439 AAAGTTTCACTGACCTTCAAAATTAAGACAAATTAATTAACATCATGATGATCT 498  
DB 361 CAAGCGTTTCTCAAAAATGATGTTTAATTAAGCATTAATGACATACCGGTGAAGAAG 420  
QY 499 CCGTGTATCATATGATCATCTCGTGTGTTTAAAAAATTTCCAAACCTTGTCAGATTA 558  
DB 421 C---CGTTAAAGTGTCTAAGTGTATGTCAGCGTTTATTCGAAAAACCATTTAAAAAGC 477  
QY 559 CATTTTCATATTAACCACTGACCAAAATTTAAATGTCCTAAATGTGTAAAGGTGT 618  
DB 478 CACTCAAGACACATACAGCCAAAAGCCGTATTAATGTATGATGATGATCAAGCGTTT 537  
QY 619 GTTGGAAAAAAGTTTATCTTCAATATGTTAAGTATGATGATTTTCAATGATCAAA 678  
DB 538 TCTCAGAAATCAAAATTTAAACAAGCATTAACAAGGACACATGCGGTGAAGAAGCTGTTAA 597  
QY 679 ATATGACCTTGTGATTA 695  
DB 598 TCGCATTAAGTGTATTA 614

RESULT 5  
AK032220  
LOCUS AK032220 1609 bp mRNA linear HTC 03-APR-2004  
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
homolog [Mus musculus], clone:6430503015 product:MSZF33 (FRAGMENT)  
ACCESSION AK032220  
VERSION AK032220.1 GI:26328050  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Nomura, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akhiba, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
location/Qualifiers  
1. 1609  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM:DB:6430503015"  
/db\_xref="taxon:10090"  
/clone="6430503015"  
/sex="male"

## CDS

```

/cisue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
330..1223
/note="unnamed protein product; MSZF3 (FRAGMENT) homolog
(mus musculus) (SPTP088251, evidence: FASTA, 87.1%ID,
100%length, match=255)
putative"
/codon_start=1
/protein_id="BAC27766.1"
/db_xref="GI:26328051"
/translation="MLETYRNLAAGVGNMEDSNIEBHCESSRRHGHNRHTGEKPYE
GIQGEAFVHSSILQMKIITHEKRYKCNQCKAVSRHSILQIHKRTSGEKPYCN
OGKAFQSHSLKIHMTHTGKPYKCNQCKAFACNKLQKERTHTGKPYKCNQ
NKAPEVSYQVHKHTHTGKPYKCNQCKAFARHSLKTKYHKITHTGKPYKCNQCK
ALAHSTLQVHQHTHTGKPYECCQCKAPANOSYFOVHKRHTGKPYKCNQCKAF
VSSDLRHRHVRHTTLOW"
1594..1599
/note="putative"
1609
/note="putative"

```

## ORIGIN

```

Query Match      7.9%; Score 98.2; DB 3; Length 1609;
Best Local Similarity 51.2%; Pred. No. 3.6e-09;
Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

```

```

QY 60 ACCTCCCAAAAGTATATTTGCACATATGAGGGGTGTGATTAAGCTTAATATGACCATC 119
DB 521 ACATACTGAGAGAAACGTACAAATCAATCAATGATGATTAAGCTTAATCAGACACAG 580
QY 120 ATTATAGACCACTTAAAGAACCCAGTAATGATGACGACCGTAATTAAGTACAGGGA 179
DB 581 CATTCTACAAATACATTAAGAAACACATGATGAGAGAAACCTTAATGATGTA-----A 634
QY 180 CGATTGTGATTAAGCATTTTTCAGAAATATCATATTTGGAACACATATGATCATCTC 239
DB 635 TCATGTGTGATTAAGCTTTTACACAAACACATGATCTCAAAATATCATATGATGATCAT 694
QY 240 CGAAAAAAACCATTCATGTTTCAAGTGTGTGTAAGGGGTTAATTCGACACACCTT 299
DB 695 TGGAGAGAAACCTTACAAATGATATCAATGTGTGTAAGCTTTGATGATCATATAAAT 754
QY 300 GAAAAAGCATGAATTCACCATACAAAGTCAATTAATGATACATTTGAAAATTTGTCAAGA 359
DB 755 TCAAAAAACATGAAGAACACATCTGAGAGAAACCTTAACAAGTATCATATGATTA 814
QY 360 AGCATTT---TATAACATCAATCTTTAAGACATCATATATATCTGTCATGAAAAAAC 416
DB 815 AGCCTTTGTATATGAAGTTATTTACAAAGTTCAATTAAGAAACACATCTGAGAGAAAC 874
QY 417 ATTAACGTGAACCAATGATATAAAGTTTTCATCGACCTTCAAAATTAAGACACAACTAA 476
DB 875 TTACAATGATATGAATGTGTGTAAGCTTTGACGACGACATCATCTCAAAAGTCAATA 934
QY 477 ATTAACCATCATGTGTGATCTCTGCTTATCAATGATCATCTGTTGTTTAAAAA 536
DB 935 AATAACCATATCTGAGAGAAAC---CTTACAAATGATATCAATGTGTGTAAGCTTTCG 991
QY 537 TTTTCAAACTTGCAGATATTACAAATTTCAATATAAACAACCTGCATCCAAACCTTAATG 596
DB 992 ATATCATATGATCACTCCAGTATCATCAAGAACCATATCTGAGAGAAAGCCCTATGATG 1051
QY 597 TCCTAAATGTGTGAAGAGTTGTGTGGAAAAAAAGTTTATCTTACATATGTTAACTCA 656
DB 1052 TGACCAATGTGTGCAAGGCTTTGCAAAATCAAAAGTTATTTCCAAAGTACATTAAGAAATACA 1111
QY 657 TGATG 661
DB 1112 TACTG 1116

```

RESULT 6

BC047646  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL

REMARK  
COMMENT

BC047646 2647 bp mRNA linear HTC 19-NOV-2003  
Homo sapiens zinc finger protein 85 (HPF4, HNF1), mRNA (cDNA clone  
IMAGE:4828290), with apparent retained intron.  
BC047646  
GI:29126804  
HTC.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
1 (bases 1 to 2647)  
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loguelfano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2647)  
Strausberg, R.  
Direct Submission  
Submitted (03-MAR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomes Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAC Project, 90 Row: 1 Column: 7  
This clone has the following problem: retained intron.  
location/Qualifiers

FEATURES  
source  
1..2647  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4828290"  
/issue\_type="Testis"  
/clone\_lib="NIH MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

## ORIGIN

Query Match

7.9%; Score 97.8; DB 3; Length 2647;

Best Local Similarity 50.1%; Pred. No. 4.5e-09; Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;

QY 7 GAAAGTACGAAACCAATCGATATCATCTTTAATATCTCTCTCTTCATCAGCTCC 66  
Db 1235 GAATGTGGCAAACTTTTAAACGATCTCTCACTCTACTACCATTAAGATTAATCTACT 1294  
QY 67 AAAAGTATTTTGCACATATGAAAGGTGTATTAAGCCTTAAATGACCATTAATTA 126  
Db 1295 GGAGAGAAACCCATCAAAATGTAAAGATGTGTAAGCCTTTTAAACGATCTCAACCTT 1354  
QY 127 GAGCAACATTTAAGAACCACTAATGATCGACCTTAATTAATGATACAGGAGCATTTGT 186  
Db 1355 ACTACCCATGAAAAATTCATCTGAGAGAAACCTTACAAAGT-----GAAAGATGT 1408  
QY 187 GATAAAGCATTTTTCAGAAAAATCACATTTGGAACAATATTTGATCATCTCCGAAAA 246  
Db 1409 GCGAAAGCCTTTAAGAGAGTCCCAACCTTACATCAATAGATTAATCTACTGAGAG 1468  
QY 247 AAACATTCATTTTCACTGTGTGTGTAAGGGGTAAATTTCTGACAACTTGAAAGA 306  
Db 1469 AAACCTACAAATGTAAAAATGTGAAAAAGCCTTTAACAGTCTGACACCTTACACA 1528  
QY 307 CATGAATACCCATCAAAAGTATTAATGTACATTTGAAATTTGTCAAGAGCATTT 366  
Db 1529 CATGAGATTAATCTACTGAGAGAAACCCATCAAAATGTGAAAAATGTGAAAAAGCCTTT 1588  
QY 367 TATPAA---ACATCAATCTTTAAGACATCATATATATCTGTTATGAAAAAACATTAAG 423  
Db 1589 AATCATTTTCTCACACCTTACTACATTAATTAATCTACTGAGAGAAACCTTACAAA 1648  
QY 424 TGTAAACATGTAATAAGTTTCACTGACCTTCAAAATTAAGACAACTTAATTAATA 483  
Db 1649 TGTAAAGAAATGTGTAAAGCTTTTAAACATCTTTCAACCTTACTAAACATTAAGATTA 1708  
QY 484 CATCATGTGTGATCTCTGCTTATCATGTGATCATCTGCTGTTTAAAAATTTCCAA 543  
Db 1709 CATACGTGAGAGAGAGCTTACAAATGTAAAGATGTGAAAAAGCCTTTTAAACCAATCTCA 1768  
QY 544 ACTTGTCAGTATTAATTAATTTATATTAATAAACAATGCAATCAAACTTAATGCTTAA 603  
Db 1769 AAACCTACTGAAACATTAAGAAATTTCTACTGAGAGAAACCTTA---TGAATGTGAAAA 1825  
QY 604 TGTGTAAAGTGTGTGTGGAATAAGTTTATCTTCAATATGTTAAATGTCATGATGA 662  
Db 1826 TGTGGCAAAAGCTTTTAAACAGTCTCTCAAAATCTTACTAGACATTAAGAAAAAGTCAATACA 1884

RESULT 7  
AUI23448 745 bp mRNA linear EST 01-AUG-2002  
LOCUS AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA  
DEFINITION sequence.

ACCESSION AUI23448  
VERSION AUI23448  
KEYWORDS GI:10948164  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 745)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.

TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,  
Suzuki,Y., Sugano,S., Isogai,T.)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yama, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975

Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES

source  
1..745  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RM2000319"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/clone\_id="NT2RM2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

ORIGIN

Query Match 7.9%; Score 97.6; DB 1; Length 745;  
Best Local Similarity 50.2%; Pred. No. 4.5e-09;  
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

QY 7 GAAAGTACGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCAGCTCC 66  
Db 87 GAATGTGGCAAACTTTTAAACGATCTCACTTACTACCATTAAGATTAATCTACT 146  
QY 67 AAAAGTATTTTGCACATATGAAAGGTGTATTAAGCCTTAAATGACCATTAATTA 126  
Db 147 GAGAGAAACCCATCAAAATGTAAAGATGTGTAAGCCTTTTAAACGATCTTCAACCTT 206  
QY 127 GAGCAACATTTAAGAACCCAGTAATGATCGACCTTAATTAATGATACAGGAGCATTTGT 186  
Db 207 ACTACCCATGAAAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAAGATGT 260  
QY 187 GATAAAGCATTTTTCAGAAAAATCACATTTGGAACAATATTTGATCATCTCCGAAAA 246  
Db 261 GCGAAAGCCTTTAAGAGAGTCCCAACCTTACTACATTAAGATTAATCTACTGAGAG 320  
QY 247 AAACATTCATTTTCACTGTGTGTGTAAGGGGTAAATTTCTGACAACTTGAAAGA 306  
Db 321 AAACCTACAAATGTAAAAATGTGAAAAAGCCTTTTAAACAGTCTGACACCTTACTACA 380  
QY 307 CATGAATACCCATCAAAAGTATTAATGTATTAATGTAATAATTTGTCAAGAGCATTT 366  
Db 381 CATGAGTAAATCTACTGAGAGAGAAACCCATCAAAATGTGAAAAATGTGAAAAAGCCTTT 440  
QY 367 TATPAA---ACATCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTAAG 423  
Db 441 AATCATTTTCTCACACCTTACTACATTAAGATTAATCTACTGAGAGAAACCTTACAAA 500  
QY 424 TGTAAACATGTAATAAGTTTCACTGACCTTCAAAATTAAGCAATCAATTAATAA 483  
Db 501 TGTAAAGAAATGTGTGTAAGCCTTTTAAACATCTTCAACCTTACTTAAACATTAAGATTA 560  
QY 484 CATCATGTGTGATCTCTGCTTATCAATGATGATCATCTGCTGTTTAAAAATTTCCAA 543  
Db 561 CATACTGAGAGAAACCTTACAAATGTAAGATGTGAAAAAGCCTTTTAAACCAATCTCA 620  
QY 544 ACTTGTCAGTATTAATTAATTTCAATTAATAAACAATGCAATCCAAAATTAATGCTTAA 603  
Db 621 AAACCTACTGAAACATTAAGAAATTTCACTCTGAGAGAAACCTTA---TGAATGTGAAAA 677  
QY 604 TGTGTAAAGTGTGTGTGGAATAAGTTTATCTTCAATATGTTAAATGTCATGATGA 662  
Db 678 TGTGGCAAAAGCTTTTAAACAGTCTCTCAAAATCTTACTAGACATTAAGAAAAAGTCA 731

RESULT 8  
CO434402 727 bp mRNA linear EST 06-JUL-2004  
LOCUS UI-M-HX0-csl-h-09-0-UI.r1 NIH\_BMAP\_HX0 Mus musculus cDNA clone  
DEFINITION IMAGE:30687440 5', mRNA sequence.

ACCESSION CO434402  
 VERSION CO434402.1 GI:49680696  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 727)  
 NIH-MGC <http://mgs.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
 Tissue Procurement: Dr. James Lin University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/mouseefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
 Seq primer: pyx-5.  
 Location/Qualifiers  
 1..727  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30687440"  
 /issue\_type="whole eye"  
 /dev\_stage="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_1lb="NIM\_BMAP\_HX0"  
 /note="Organ: Eye; Vector: pyx-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATTAATTCG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN  
 Query Match 7.5%; Score 93.4; DB 7; Length 727;  
 Best Local Similarity 52.1%; Pred. No. 3e-08;  
 Matches 261; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

4 AGTGAAGAGCAAGCAACCAATGATATCATCTTTAATCTCTCTTATCATACGT 63  
 DB AATCACTGTGATTAAGCTTTTCTCGATACAGTACTTCAATCTCAATGAGAACAACAT 137  
 64 CCCAAAAGATATTTTGCAATATGAAAGGTGTGATAAGCCATATATGACCATATTA 123  
 DB ACTGAGAGAAACCTTCAATGTATACGTGTGATTAAGCCTTTCTGAGAAATGTAGT 197  
 124 TTAGAGCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTATACGTGACAT 183  
 DB CTCCAAAACCATGTAGATGATACACTGAGAGAAACCCCTACAAATGTA-----ATCAA 251  
 184 TGTGATTAAGCATTTTTCAGAAATACATTTTGAACAATATTTGATTCATTCGGA 243  
 DB TGTGATTAAGCCCTTTTCCATTAATCATCTCCACATTTATAGAAAGACATATCCGAG 311  
 244 AAAAAACATTCATTTTCAGTGTGTGTGTAAGGGGTTAATTTCTGACAAACATTTGAA 303  
 DB GAAAAACCCATGAAGATATATCATGTGTGATTAAGCCTTTCCGATTAACATATCACTCC 371

304 AGACATGAATCACCACATACAAAGTCATTTAATGATCATTTGAATTTGCAGAAGCA 363  
 DB ATTCATAGAGAAACACATATCTGAGAGAAAGCCTTACAAATGTATATCATGTGTATAAGCC 431  
 364 TTTTATTAACATCAATCTTTAAGACATCATATATATC---TGTTCATGAAAAAACATTA 420  
 DB TTTTTCATATACCTTTTCTCTTCAAACTGACAGAAAGACATATCTGAGAGAAACCATTC 491  
 421 ACGTGAACATGTATTAAGTTTCACTGCACCTTCAAAATTTAGACAAATTAATTA 480  
 DB AATGTAAATCAATGATTAAGTCTTTCTCAATACAGTATCTTCAACTATAGAGA 551  
 481 AAACATCAATGATGATCTCCT 501  
 DB ACACATATCTGAGAGAAACCT 572

RESULT 9  
 LOCUS CN352904 631 bp mRNA linear EST 16-MAY-2004  
 DEFINITION 17000600057466 GRN\_PREHEP Homo sapiens cDNA 5', mRNA sequence.  
 ACCESSION CN352904  
 VERSION CN352904.1 GI:47352838  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 631)  
 Branderberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebkoweki, J. and Stanton, L.W.  
 Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation  
 Nat. Biotechnol. 22 (6), 707-716 (2004)  
 Contact: Branderberger R  
 Regenerative Medicine  
 Genon Corporation  
 230 Constitution Drive, Menlo Park, CA 94025, USA  
 Tel: 650 473 8658  
 Fax: 650 473 7760  
 Email: [rbranderberger@genon.com](mailto:rbranderberger@genon.com)  
 Insert Length: 631. Std Error: 0.00.  
 Location/Qualifiers  
 1..631  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /issue\_type="embryonic stem cells, DMSO-treated H9 cell line"  
 /clone\_1lb="GRN\_PREHEP"  
 /note="oligo dT primed, full-length enriched cDNA library from DMSO-treated hES cell line H9 (p22) maintained in feeder-free conditions"

ORIGIN  
 Query Match 7.5%; Score 92.4; DB 7; Length 631;  
 Best Local Similarity 54.1%; Pred. No. 4.7e-08;  
 Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

69 AAAGTATATTTGACATATGAAGGGTGTGATAAGCCCTAATCGACCATATTAATA 128  
 DB AAGAGAGCCCTTCAAAATGTAAAGTAAGAAAGCTTTTAACCAATCTCAAAACTTAC 142  
 129 GCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGAAGATTTGA 188  
 DB TGAACATTAAGAAATTCATCTGAGAGAAACCTATGAATGT-----GAAAAATGTGG 196  
 189 TAAAGCATTTTTCAGAAAAATCAATTTGAAAAACATATTTGTATCATTTCCGAAAAAAA 248  
 DB CAAAGCTTTTAACCAAGTCTCAAAATCTTACTGACATTAAGAAAAAGTCATACAGAAAGAA 256

QY 249 ACCATTCCATTGTCAGTGTGTGAAGGGGTTAATTCGACCAACCTGTAAGACA 308  
DB 257 ACCCTTACAATGTGAAGAAATGGCAAGTTTAAATGGCCCTCAACCTTACTATCA 316  
QY 309 TGAATACCCACATGCAAGTCAATTTAAATGTCATTTGAAGTGAAGAACATTTA 368  
DB 317 TAAAGATTAATTCATGAGAGAGAACCAATGAAATGAGATGGCAAGCTTTAA 376  
QY 369 TAAATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAAATTAACGTG 425  
DB 377 CCAATCTCAAACTTACCAACATTAAGAAATTCATCTGAGAGAAACCTACACATG 436  
QY 426 TAAACATGATTAAGATTTCACTGACCTTCAAAATTAAGCAACATTAATTAACA 485  
DB 437 TGAAGATGTGGCAAGCCCTTAACCAAGTCTCAAACTTACTTAACATTAAGATTC 496  
QY 486 TCATGTGTGATCTCT 501  
DB 497 TACTGAGAAAAAACCCT 512

RESULT 10  
BC036394 2672 bp mRNA linear HTC 19-NOV-2003  
LOCUS Homo sapiens zinc finger protein 85 (HPF4, HTP1), mRNA (cDNA clone  
IMAGE:5259399), containing frame-shift errors.  
ACCESSION BC036394  
VERSION BC036394.1 GI:23025784  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 2672)  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschuler, S.F., Zeeberg, B., Buetow, K.H., Scheaffer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stalderon, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Adams, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Wyllie, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Sanchez, A., Whiting, M., Madan, A., Madan, A., Rodriguez, S.,  
Boutard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.B.,  
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE 2 (bases 1 to 2672)  
JOURNAL Strausberg, R.  
MEDLINE Direct Submission  
PUBMED Submitted (05-AUG-2002) National Institutes of Health, Mammalian  
22388257 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
JOURNAL Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590,  
USA  
AUTHORS NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiroki  
Tohliki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>

contact: [amadane@systembiology.org](mailto:amadane@systembiology.org)  
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IPAK Plate: 73 Row: a Column: 16  
This clone has the following problem: frame shifted.  
Location/Qualifiers  
1..2672  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5259399"  
/issue\_type="Brain, hippocampus"  
/clone\_id="NIH\_MGC\_95"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

ORIGIN  
Query Match 7.5%; Score 92.4; DB 3; Length 2672;  
Best Local Similarity 54.1%; Pred. No. 5.3e-08;  
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;  
QY 69 AAGATTAATTTGACATATGAAGGGTGTGATTAAGCCTATTAATGACCATCATTAATGA 128  
DB 1745 AGAAGAGCCTTAACAATGTAAAGATGTGAAGAAAGCTTTTAACCAATCCTCAAACTTAC 1804  
QY 129 GCAACATTTAAGAACCCACAGTAATGATCGAACCGTATTAATGTACAGTGACGATTTGGA 188  
DB 1805 TGAACATTAAGAAATTCATCTGAGAGAAACCCATTAATGT-----GAAAAATGTGG 1858  
QY 189 TAAACATTTTTCAGAAATCAATTTGGAACACATATTGTATCATCTTCGAAAAA 248  
DB 1859 CAAGCTTTTAACCAAGTCTCAATCTTACATGACATTAAGAAATCATACGAAAGAA 1918  
QY 249 ACCATTCCATTGTCAGTGTGTGAAGGGGTTAATTCGACCAACCTGTAAGACA 308  
DB 1919 ACCCTTACAATGTGAAGAAATGGCAAGTTTAAATGGCCCTCAACCTTACTATCA 1978  
QY 309 TGAATCAACCATCAAGTCAATTAATGTATGATTTGAAGAAATTCGAAAGCAATTTTA 368  
DB 1979 TAAAGTAAATTCATATCTGAGAGAAACCAATACAAATGTGAAGATGTGCAAGCTTTTAA 2038  
QY 369 TAAATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAAATTAACGTG 425  
DB 2039 CCAATCTCAAACTTACCAACATTAAGAAATTCATCTGAGAGAAACCTACACATG 2098  
QY 426 TAAACATGATTAAGATTTTCACTGACCTTCAAAATTAAGCAACATTAATTAACA 485  
DB 2099 TGAAGATGTGGCAAGCCCTTAACCAAGTCTCAAACTTACTTAACATTAAGAAATTC 2158  
QY 486 TCATGTGTGATCTCT 501  
DB 2159 TACTGAGAAAAAACCCT 2174

RESULT 11  
A0005136 591 bp DNA linear GSS 27-JUN-1998  
LOCUS A0005136  
DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017,  
genomic survey sequence.  
ACCESSION A0005136  
VERSION A0005136.1 GI:3082581  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 591)  
AUTHORS Adams, M.D., Rounseley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,  
Golden, K., Berry, K., Granger, D., Sun, E., Wible, C., Shizuya, H.,  
Simon, M. and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: CIT-HSP-2290017.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@igf.org  
Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
http://www.igf.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
source

1..591  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7150844"  
/db\_xref="taxon:9606"  
/clone="2290017"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 7.4%; Score 92; DB 8; Length 591;  
Best Local Similarity 52.2%; Pred. No. 5.7e-08;  
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

7 GAAAGTGCAGAAACCAATGCATATCATCTTTATATCTTCTTCTTCATCAGTCCC 66  
69 GAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACATACATAGATTAATCACT 128  
67 AAAAAGTATTTTGCATATGAAAGGTGTGATTAAGCCTTATATGACCATCATTA 126  
129 GGAGAAAGCCCTTCAATGTGAAGATGTGTAAGCCTTTTAACACCCCTTCAGCCCTT 188  
127 GAGCAATTTAAGAACCCAGTAATGATCGACCGTATTAATGTAAGTGAAGATTTG 186  
189 ACTACACATTAAGTTCATCTGTTAAAGAAAAACCTTCAAAATGT-----GAAGATGT 242  
187 GATTAAGCATTTTTCAGAAAAATCACAATTGGAACAATTTGTATCACAATTCGAAAA 246  
243 GACAAAGCTTTTAACGATTTCTATACCTTAACATTAAGATTAATTCATTCGAGAG 302  
247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTATTTCTGACAAACCTTGAAGA 306  
303 AATCTTCAAAAGTGAACATGTGGAAAGGCTTTAACTGCTTCACACCTTACAAA 362  
307 CATGAATCAACCCATCAAAAGTCATTTAATGTATCATTTGAAATTTGTCAAGAGATTT 366  
363 CATTAAGATTAATTAATCTGAGAGAAACCTTACAAATGTAAAGAAATGTGCAAGCCTTT 422  
367 TATTAACAT--CAATCTTTAAGACATCATATATATCTGTCATGAAAAAACATTAACG 423  
423 AATGTGCTTACACCTTACTACATTAAGATGATTCATCTGAGAGAAACCTTACAA 482  
424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTTAGCAACAATTAATTA 483  
483 TGTGAAGAAATGTGCAAGCCTTTAACAACCTCTCAAACTTACTATACATAAGATTA 542  
484 CATCATGCTG 493  
543 CATCTGAG 552

RESULT 12  
CN362485

LOCUS CN362485 687 bp mRNA linear EST 16-MAY-2004  
DEFINITION 17000424188118 GRN\_EB Homo sapiens cDNA 5', mRNA sequence.  
ACCESSION CN362485  
VERSION CN362485.1 GI:47362419  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1..687  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cissue\_type="embryonic stem cells, embryoid bodies  
derived from H1, H7 and H9 cells"  
/clone\_lib="GRN\_EB"  
/note="Toligo dT primed, full-length enriched cDNA library  
from embryoid body outgrowth derived from hES cell lines  
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free  
conditions."

## ORIGIN

Query Match 7.4%; Score 92; DB 7; Length 687;  
Best Local Similarity 52.2%; Pred. No. 5.7e-08;  
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

7 GAAAGTGCAGAAACCAATGCATATCATCTTTATATCTTCTTCTTCATCAGTCCC 66  
69 GAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACATACATAGATTAATCACT 128  
67 AAAAAGTATTTTGCATATGAAAGGTGTGATTAAGCCTTATATGACCATCATTA 126  
183 AGATGTGCAGAAACATTTTGCATGCTTTTCACACCTTACATACATAGATTAATCA 242  
243 AGATGAATTTTTCAGAAATGTGAAGCATATGAAAGCCTTTAACTGCTTCACACCTT 302  
127 GAGCAATTTAAGAACCCAGTAAATGATCGACCGTATTAATGTACAGTGAAGATTTG 186  
303 AATTAACATTAAGAAATTAATCTGAGAAAAACCTTACAAATGTAAA-----GATGT 356  
187 GATTAAGCATTTTTCAGAAAAATCACAATTGGAACAATTTGTATCACAATTCGAAAA 246  
357 GGCAAGCCTTTTACAGACCTTACACCTTATTAAGCATTAAGAAATTTCTATCTGAAG 416  
247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGCAACAACCTTGAAGA 306  
417 AAACCTTCAAAATGTGAAGATGTGCAAGCCTTTAAACAGATCAAGACCTTATCA 476  
307 CATGAATCAACCCATCAAAAGTCATTTAATGTACATTTGAAATTTGTCAAGAGATTT 366  
477 CATTAATTAATTAATCTGAGGAAATTTCCCTCAAAATGTGAGAAATGTGTAGAGCTTTT 536  
367 TATTAACATCAATCTTTAAGACATCATATATATCTGTT-----CATGAAAAACATTAAG 423  
537 AACCAAGCTTCAAGCTTACTGAAACATTAATTAATTCATACCGAGAGAAACCTTATGA 596  
424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTTAGCAACAATTAATTA 483  
597 TGTGAAGAAATGTGCAAGCCTTTTAAACGATCTCTCAAACTTACTGAAATTAAGTACAT 656

QY 484 CATCATGTG 493  
 DB 657 CATACTGAG 666

RESULT 13  
 CN417582 752 bp mRNA linear EST 16-MAY-2004  
 LOCUS 17000424010604 GRN\_EB Homo sapiens cDNA 5', mRNA sequence.  
 DEFINITION CN417582  
 ACCESSION CN417582  
 VERSION CN417582.1 GI:47405176  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 752)  
 Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G.J., Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebowicki, J. and Stanton, L.W.  
 Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation  
 Nat. Biotechnol. 22 (6), 707-716 (2004)  
 CONTACT: Brandenberger R  
 JOURNAL  
 COMMENT  
 Reenerative Medicine  
 Gen Corporation  
 230 Constitution Drive, Menlo Park, CA 94025, USA  
 Tel: 650 473 8658  
 Fax: 650 473 7760  
 Email: rbrandenberger@genon.com  
 Insert Length: 752 Std Error: 0.00.  
 Location/Qualifiers  
 1..752  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /isue\_type="embryonic stem cells, embryoid bodies  
 derived from H1, H7 and H9 cells"  
 /clone\_lib="GRN EB"  
 /note="oligo dt primed, full-length enriched cDNA library from embryoid body outgrowth derived from hbs cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions."

ORIGIN  
 Query Match 7.4%; Score 92; DB 7; Length 752;  
 Best Local Similarity 52.0%; Pred. No. 5.8e-08;  
 Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATGATGATCATCTTTAATATCTTCTTCATCAGCGTCC 66  
 DB 72 GAATGGGAAAGCCTTACCATCTTTCACACCTTACACATAGAATATTCATCT 131  
 QY 67 AAAAGTATATTGACATATGAAGGGGTGATAAAGCCTATTAATGACCATATTATTA 126  
 DB 132 GGAGAGAAAGCCCTTCAATGATGAAGATGTGTAAGCTTTAAACACCCCTCAGCCCT 191  
 QY 127 GAGCAACATTTAAGAACCCAGATATGATGACCCCTATTAATATGATGACGATTT 186  
 DB 192 GCTACACATAGTCTTCAATGATTAAGAAAAACCTACAAATGT-----GAAAGATGT 245  
 QY 187 GATTAAGCATTTTTCAGAAATCACAATTTGAAACATATTGTATCATATTCGGAATA 246  
 DB 246 GACAAAGCTTTTAAACGATTTCTATCTTACATTAACATTAATATTCATTTCTGAGAG 305  
 QY 247 AAACATTCATTTTCACTGTGTGTGTAAGGGGTTAATCTCGACACACTTGAATA 306  
 DB 306 AATCTTCAAAATGTGAACAAATGTGCAAGGCTTTAACTGCTTCAACCCCTTACAAA 365  
 QY 307 CATGAATACCATATCAAGATCATTTAATATGATGATTTGAAATATGTCAAGAGATTT 366  
 DB 366 CATAGAAATTCATCTGAGAGAAACCTCAAAATGTGAAGATGTGCAAGGCTTT 425

QY 367 TATTAACAT---CAATCTTAAAGACATATATATATCTGTTACATGAAAAACATTAACG 423  
 DB 426 AATGTCTTACACACCTTACTACATAGATGATTAATCTACTGAGAGAAACCTTACAAA 485  
 QY 424 TGTAAACATGTATTAAGATTTCCTGACCTTCAAAATTTAGCAACATTAATTTAAA 483  
 DB 486 TGTGAAGATGTGGCAAGGCCCTTTAACACCTCTCAAAACCTATATCATTAAGATTAAT 545  
 QY 484 CATCATGTTGATCTCTCT 501  
 DB 546 CATCTGAGAGAAACCT 563

RESULT 14  
 B99387 593 bp DNA linear GSS 26-JUN-1998  
 LOCUS CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,  
 DEFINITION genomic survey sequence.  
 B99387  
 ACCESSION B99387  
 VERSION B99387.1 GI:3027197  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 593)  
 Adams, M.D., Rounley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Simon, M., and Venter, J.C.  
 Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)  
 Unpublished (1998)  
 CONTACT: Mark Adams  
 JOURNAL  
 COMMENT  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@reagen.com). BAC  
 end search page:  
 http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: M13-21;  
 Class: BAC ends.  
 Location/Qualifiers  
 1..593  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7147044"  
 /db\_xref="taxon:9606"  
 /clone="2281A10"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /clone\_lib="CIT-HSP"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

ORIGIN  
 Query Match 7.3%; Score 91; DB 8; Length 593;  
 Best Local Similarity 54.4%; Pred. No. 8.9e-08;  
 Matches 231; Conservative 0; Mismatches 185; Indels 9; Gaps 2;

QY 69 AAAGTATTTTGAATGAAGGGGTGATTAAGCCTATTAATGACATCATTTATTA 128  
 DB 47 AAAGAAACCTTCAAAATGTGAAGATGTGGCAAGCTTTAGCCAGTCTCAACCTTAG 106  
 QY 129 GCAACATTTAAGAACCCAGATATGATGACCCGATTAATGATGATGACGATTTGA 188  
 DB 107 AAAACATAGATTAATTCATCTGAGAGAAACCTTACAAATGT-----GAAAGATGTG 160  
 QY 189 TAAACATTTTTCAGAAATATCATTTGAAACACATATTTGATCATTCGGAATAA 248





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:03:15 ; Search time 73 Seconds  
(without alignments)  
2182.812 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229  
Sequence: 1 MESSDETKSISLSISSSSSS.....PLVKARMDLPLETSVISR 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	2229	100.0	412	3	AA93316 A transcr
2	2225	99.8	412	3	ABP73656
3	445	20.0	554	6	ABJ25655 Aspergill
4	445	20.0	554	6	ABJ26255 Aspergill
5	416.5	18.7	1212	4	ABG00399 Novel hum
6	407	18.3	664	8	ADN9509 Novel hum
7	403	18.1	1230	4	AAU30831
8	400	17.9	404	7	ADMO5916
9	400	17.9	803	8	ADN06034
10	400	17.9	809	4	AAW38689
11	399	17.9	444	8	ADN9691 Novel hum
12	398	17.9	444	8	ADN9690 Novel hum
13	398	17.9	487	8	ABO60215 Human gen
14	398	17.9	553	6	ABU96690 Human nuc
15	398	17.9	553	6	ADU60246 Secreted
16	397.5	17.8	423	8	ADQ29657
17	397.5	17.8	423	8	ADP24748 PRO polyp
18	395.5	17.7	365	3	AA93317 A human t
19	395.5	17.7	409	4	AA975181
20	395.5	17.7	423	2	AA91305
21	395.5	17.7	1191	7	ADJ70299 Human hea
22	395	17.7	622	4	AAW78947 Human pro
23	395	17.7	631	4	AAW79931 Human pro
24	395	17.7	632	4	ABG18386 Novel hum
25	395	17.7	719	4	ABG16953 Novel hum

26	393.5	17.7	1191	8	ADN04208	Adn04208 Antipeori
27	393.5	17.7	1472	4	AAU31578	Aau31578 Novel hum
28	393	17.6	517	4	ABG03375	Abg03375 Novel hum
29	392.5	17.6	1050	4	ABG09685	Abg09685 Novel hum
30	392.5	17.6	1050	7	ADC32985	Adc32985 Human nov
31	392	17.6	555	6	ADA54763	Ada54763 Human pro
32	391	17.5	592	7	ADM04984	Adm04984 Human pro
33	390.5	17.5	878	4	AAU29528	Aau29528 Novel hum
34	390.5	17.5	1520	4	ABG27130	Abg27130 Novel hum
35	390	17.5	540	7	AD655594	Ad655594 Human pro
36	390	17.5	576	6	ADA54797	Ada54797 Human pro
37	390	17.5	620	6	ABU96725	Abu96725 Human nuc
38	389	17.5	891	8	ADQ67674	Adq67674 Novel hum
39	388.5	17.4	501	4	ABR71369	Abb71369 Drosophi1
40	388	17.4	568	5	ABR79872	Abb79872 TRAF6-1th
41	387.5	17.4	723	8	ADN9696	Adn9696 Novel hum
42	387.5	17.4	742	8	ABO58351	Abos8351 Human gen
43	387.5	17.4	839	7	ADJ70044	Adj70044 Human hea
44	387.5	17.4	872	7	ADC37535	Adc37535 Human nuc
45	386.5	17.3	327	8	ADR08458	Adr08458 Human pro

## ALIGNMENTS

RESULT 1	AA93316	standard; protein; 412 AA.
ID	AA93316	
XX	AC	AA93316;
XX	DT	04-SEP-2000 (first entry)
XX	DE	A transcription factor designated CATP11A.
XX	KM	Transcription factor; CATP11A; DNA-binding protein;
KW	ribosomal RNA 5S gene; fungal infection.	
XX	OS	Candida albicans.
XX	FH	Key Location/Qualifiers
FT	Misc-difference 193	/note= "Ser encoded by CTG"
FT	Misc-difference 339	/note= "Ser encoded by CTG"
FT		
XX	PN	WO200028037-A1.
XX	PD	18-MAY-2000.
XX	PF	09-NOV-1999; 99WO-FR002739.
XX	PR	10-NOV-1998; 98FR-00014147.
XX	PA	(HMRI ) HOECHST MARION ROUSSEL.
XX	PI	Bordon-Pallier F, Camier S, Sentenac A;
XX	DR	WPI; 2000-376549/32.
XX	DR	N-PSDB; AAA15398.
XX	PT	New nucleic acid encoding Candida albicans transcription factor, useful
PT	e.g. in screening for antimycotic agents and for immunization.	
XX	XX	Claim 12; Page 35-36; 45pp; French.
XX	XX	The present sequence represents a Candida albicans transcription factor,
CC	designated CATP11A. The polypeptide is a DNA-binding protein, which is	
CC	involved in initiating transcription of the ribosomal RNA 5S gene. The	
CC	polynucleotide is used to screen for its specific inhibitors, potentially	
CC	useful as antimycotic agents, to raise an antibody response that is	
CC	protective against fungal infection and to raise antibodies. Such	
CC	antibodies, as well as the polypeptides and polynucleotides are used in	

CC Compositions for diagnosing and treating fungal infections, e.g. by  
CC detecting polymorphisms and mutations

XX Sequence 412 AA;

Query Match 100.0%; Score 2229; DB 3; Length 412;  
Best Local Similarity 100.0%; Pred. No.1e-167;  
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSESDETKSISLSISSSSSRPKKYICTYEGCDKAYNPSLLEQHLRTHSNDPRPYCTVD 60  
DB 1 MSESDETKSISLSISSSSSRPKKYICTYEGCDKAYNPSLLEQHLRTHSNDPRPYCTVD 60  
QY 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYSFKCTFENCQE 120  
DB 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYSFKCTFENCQE 120  
QY 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTRPBKLQHLKHHGSPAYQCDHPGCFKNF 180  
DB 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTRPBKLQHLKHHGSPAYQCDHPGCFKNF 180  
QY 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFPK 240  
DB 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFPK 240  
QY 241 KNELVENHNI FHDGNI PDLLKETEYVKLENI LDOGSKLNNLHELETEKLYAEDEBDE 300  
DB 241 KNELVENHNI FHDGNI PDLLKETEYVKLENI LDOGSKLNNLHELETEKLYAEDEBDE 300  
QY 301 DSLDEKSDVRSDSMSAORSIKSFTASLSEGSVSKLINSNGKINCPRNCCDRMFSREY 360  
DB 301 DSLDEKSDVRSDSMSAORSIKSFTASLSEGSVSKLINSNGKINCPRNCCDRMFSREY 360  
QY 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKAMDLLPNETSVISR 412  
DB 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKAMDLLPNETSVISR 412

RESULT 2  
ABP73656

ID ABP73656 standard; protein; 412 AA.

XX ABP73656;

DT 30-JAN-2003 (first entry)

DE Candida albicans essential protein SEQ ID NO 7493.

KM Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal.

XX Candida albicans.

OS

PN WO200253728-A2.

PD 11-JUL-2002.

PF 26-DEC-2001; 2001WO-US049486.

PR 29-DEC-2000; 2000US-0259128P.

PR 20-FEB-2001; 2001US-00792024.

PR 22-AUG-2001; 2001US-0314050P.

XX

PA (ELIT-) ELITRA PHARM INC.

PI Roemer T, Jiang B, Boone C, Busey H, Ohlsen KJ,

DR WPI; 2002-566694/60.

DR N-PSDB; AB232206.

XX

PT Constructing strains for identifying gene products as effective targets

PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.  
XX  
XX Claim 44; SEQ ID NO 7493; 167bp + Sequence Listing; English.

CC The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office

XX Sequence 412 AA;

Query Match 99.8%; Score 2225; DB 5; Length 412;  
Best Local Similarity 99.8%; Pred. No.2.2e-167;  
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSESDETKSISLSISSSSSRPKKYICTYEGCDKAYNPSLLEQHLRTHSNDPRPYCTVD 60  
DB 1 MSESDETKSISLSISSSSSRPKKYICTYEGCDKAYNPSLLEQHLRTHSNDPRPYCTVD 60  
QY 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYSFKCTFENCQE 120  
DB 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYSFKCTFENCQE 120  
QY 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTRPBKLQHLKHHGSPAYQCDHPGCFKNF 180  
DB 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTRPBKLQHLKHHGSPAYQCDHPGCFKNF 180  
QY 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFPK 240  
DB 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFPK 240  
QY 241 KNELVENHNI FHDGNI PDLLKETEYVKLENI LDOGSKLNNLHELETEKLYAEDEBDE 300  
DB 241 KNELVENHNI FHDGNI PDLLKETEYVKLENI LDOGSKLNNLHELETEKLYAEDEBDE 300  
QY 301 DSLDEKSDVRSDSMSAORSIKSFTASLSEGSVSKLINSNGKINCPRNCCDRMFSREY 360  
DB 301 DSLDEKSDVRSDSMSAORSIKSFTASLSEGSVSKLINSNGKINCPRNCCDRMFSREY 360  
QY 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKAMDLLPNETSVISR 412  
DB 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKAMDLLPNETSVISR 412

RESULT 3  
ABJ25655

ID ABJ25655 standard; protein; 564 AA.

XX ABJ25655;

DT 16-APR-2003 (first entry)

XX

DE Aspergillus fumigatus essential gene protein #313.  
 XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
 KW cancer; contamination; biofilm; antibody; immune response.  
 XX Aspergillus fumigatus.  
 OS  
 PN WO200286090-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 23-APR-2002; 2002WO-US013142.  
 XX  
 PR 23-APR-2001; 2001US-0285697P.  
 XX 27-APR-2001; 2001US-0287066P.  
 PR 05-JUN-2001; 2001US-0295890P.  
 PR 09-JUL-2001; 2001US-0303899P.  
 PR 31-AUG-2001; 2001US-0316362P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
 XX MPI; 2003-093124/08.  
 DR  
 XX  
 PT New purified or isolated nucleic acids of essential genes of Aspergillus  
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,  
 PT or for treating a non-infectious disease in a subject e.g. cancer.  
 XX  
 PS Disclosure; Page; 175pp; English.  
 XX  
 XX The invention relates to novel purified or isolated nucleic acids of  
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of  
 CC the invention are used to treat or prevent infections by a pathogenic  
 CC organism such as A. fumigatus, to treat a non-infectious disease in a  
 CC subject (e.g. cancer), to prevent or contain contamination of an object  
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a  
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for  
 CC expressing recombinant protein for characterisation, screening or  
 CC therapeutic use, as markers for host tissues in which the pathogenic  
 CC organisms invade or reside, for comparing with the DNA sequence of A.  
 CC fumigatus to identify duplicated genes or paralogues having the same or  
 CC similar biochemical activity and/or function, for comparing with DNA  
 CC sequences of other related or distant pathogenic organisms to identify  
 CC potential orthologous essential or virulence genes, for selecting and  
 CC making oligomers for attachment to a nucleic acid array for examination  
 CC of expression patterns, for raising anti-protein antibodies, as an  
 CC antigen to raise anti-DNA antibodies or to elicit another immune  
 CC response, and for identifying polynucleotides encoding the other protein  
 CC with which binding occurs or to identify inhibitors of the binding  
 CC interaction. The polypeptides may be used to raise antibodies or to  
 CC elicit immune response, as a reagent in assays designed to quantitatively  
 CC determine levels of the protein in biological fluids, as a marker for  
 CC host tissues in which pathogenic organisms invade or reside, and to  
 CC isolate correlative receptors or ligands in the case of virulence  
 CC factors. This sequence represents a protein of one of the essential genes  
 CC of Aspergillus fumigatus of the invention  
 CC  
 XX  
 XX Sequence 564 AA:  
 SQ  
 Query Match 20.0%; Score 445; DB 6; Length 564;  
 Best Local Similarity 30.3%; Pred. No. 2.3e-26;  
 Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;  
 QY 14 ISSSSSSRP---KKYICTYEGCDKAVNRPSELYSCNLTSHNDPRYKCTVDDCKAFPRKS 70  
 DB 71 ISTTSAYPSELKTHRCRPFDCGCKAFNRPARLQEHLSHNNRIPIKCTPESCKTFLRAS 130  
 QY 71 HLETHIVS-HSEKKRPHCS--VQKGVNSRQHLKRHEITL--TKSFCKT-FENCQEAFLYK 124  
 DB 131 HLNMHTKSAHTGVADYVCDRPGGSKFVTSRLRLAAHDGDKRCKTCTEYPPCNETFRK 190  
 QY 125 HGLRHHILSVH--EKLTLG-----KQCNKVFTFRPSLQAHLKHHGSGRAYQGDH- 173

DB 191 HSTLQKHWITAHILKQPFQCPHTDPSGTQCKTMAFDTAGHLRAHESRIH-TEKRFACETEC 249  
 QY 174 ----RGCFFKNPQTSWYLQHIKQSHPKLKCPKCGKGVKKGSLISM-LSHDSTM--TK 226  
 DB 250 SQHARGAENATFPPTALQIHRSVHP--QCPICALTCATSRRLRLLEVAHGVSLEERK 308  
 QY 227 IWTCDY--CDVGKFAKNELVEHYNI FHDGNI-----PDDLKETEVKLENDLDD----- 275  
 DB 309 IFPCIVPGCD-RSFTYKGNLTVHIRTWHQGEKRFVCGEFTDLSSSKVSGMN--DNGCGK 365  
 QY 276 --GSKLNLIHELETEKLVKEDEDEDEDSL-----DEKSDVSDSNQAORSIKSFASLE 329  
 DB 366 RYGSKLALBEHRIHTALGVQNAKAERQRLGTRDRQHSATATSPGSA-----LAALT 418  
 QY 330 GSKSVSKLISNGKIKNCPRNCCDRMFSEYDLRRL--KTHDDNLQRIEFL 380  
 DB 419 GEGYAE---TGRHIACTLVESCPRHFRHDYDLWMVMSGKHHSFEETRLFL 467  
 RESULT 4  
 ABJ26255  
 ID ABJ26255 standard; protein; 564 AA.  
 XX  
 XX ABJ26255;  
 AC  
 XX  
 DT 16-APR-2003 (first entry)  
 XX  
 XX Aspergillus fumigatus essential gene protein #913.  
 DE  
 XX  
 XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
 KW cancer; contamination; biofilm; antibody; immune response.  
 XX  
 OS Aspergillus fumigatus.  
 OS  
 PN WO200286090-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 23-APR-2002; 2002WO-US013142.  
 XX  
 PR 23-APR-2001; 2001US-0285697P.  
 XX 27-APR-2001; 2001US-0287066P.  
 PR 05-JUN-2001; 2001US-0295890P.  
 PR 09-JUL-2001; 2001US-0303899P.  
 PR 31-AUG-2001; 2001US-0316362P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
 XX MPI; 2003-093124/08.  
 DR  
 XX  
 PT New purified or isolated nucleic acids of essential genes of Aspergillus  
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,  
 PT or for treating a non-infectious disease in a subject e.g. cancer.  
 XX  
 PS Disclosure; Page; 175pp; English.  
 XX  
 XX The invention relates to novel purified or isolated nucleic acids of  
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of  
 CC the invention are used to treat or prevent infections by a pathogenic  
 CC organism such as A. fumigatus, to treat a non-infectious disease in a  
 CC subject (e.g. cancer), to prevent or contain contamination of an object  
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a  
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for  
 CC expressing recombinant protein for characterisation, screening or  
 CC therapeutic use, as markers for host tissues in which the pathogenic  
 CC organisms invade or reside, for comparing with the DNA sequence of A.  
 CC fumigatus to identify duplicated genes or paralogues having the same or  
 CC similar biochemical activity and/or function, for comparing with DNA  
 CC sequences of other related or distant pathogenic organisms to identify  
 CC potential orthologous essential or virulence genes, for selecting and



XX Novel human protein sequence #325.  
DE anti-inflammatory; dermatological; neuroprotective; immunomodulator;  
XX antibacterial; virucide; antiparasitic; cytostatic; gene therapy;  
KM vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;  
KM early aging; hormonal imbalance; ischemic heart disease;  
KM ulcerative colitis.  
XX Homo sapiens.  
OS  
XX WO2004038003-A2.  
XX  
XX 06-MAY-2004.  
XX  
XX 24-OCT-2003; 2003MO-US033947.  
XX  
XX 25-OCT-2002; 2002US-0421061P.  
XX 25-OCT-2002; 2002US-0421080P.  
XX 25-OCT-2002; 2002US-0421552P.  
XX 25-OCT-2002; 2002US-0421614P.  
XX 30-OCT-2002; 2002US-0422177P.  
XX 30-OCT-2002; 2002US-0422178P.  
XX 15-NOV-2002; 2002US-0426355P.  
XX 15-NOV-2002; 2002US-0426384P.  
XX 15-NOV-2002; 2002US-0426394P.  
XX 15-NOV-2002; 2002US-0426430P.  
XX 15-NOV-2002; 2002US-0426966P.  
XX 27-NOV-2002; 2002US-0429224P.  
XX 27-NOV-2002; 2002US-0429275P.  
XX 27-NOV-2002; 2002US-0429302P.  
XX 27-NOV-2002; 2002US-0429366P.  
XX 27-NOV-2002; 2002US-0429651P.  
XX 04-DEC-2002; 2002US-0430645P.  
XX 04-DEC-2002; 2002US-0430651P.  
XX 04-DEC-2002; 2002US-0430657P.  
XX 04-DEC-2002; 2002US-0430663P.  
XX 04-DEC-2002; 2002US-0430668P.  
XX 05-DEC-2002; 2002US-0430684P.  
XX 05-DEC-2002; 2002US-0430937P.  
XX 05-DEC-2002; 2002US-0430956P.  
XX 12-DEC-2002; 2002US-0431458P.  
XX 12-DEC-2002; 2002US-0432511P.  
XX 12-DEC-2002; 2002US-0433500P.  
XX 13-DEC-2002; 2002US-0433316P.  
XX 13-DEC-2002; 2002US-0433318P.  
XX 23-DEC-2002; 2002US-0436236P.  
XX 03-JAN-2003; 2003US-0437914P.  
XX 17-JAN-2003; 2003US-0440820P.  
XX 17-JAN-2003; 2003US-0440821P.  
XX 18-APR-2003; 2003US-0463700P.  
XX 18-APR-2003; 2003US-0463708P.  
XX 18-APR-2003; 2003US-0463716P.  
XX 18-APR-2003; 2003US-0463732P.  
XX 02-MAY-2003; 2003US-0467199P.  
XX 02-MAY-2003; 2003US-0467201P.  
XX 02-MAY-2003; 2003US-0467203P.  
XX 02-MAY-2003; 2003US-0467230P.  
XX 19-MAY-2003; 2003US-0471306P.  
XX 19-MAY-2003; 2003US-0471336P.  
XX 22-MAY-2003; 2003US-0472420P.  
XX 22-MAY-2003; 2003US-0472430P.  
XX 09-JUN-2003; 2003US-0476609P.  
XX 09-JUN-2003; 2003US-0476621P.  
XX 09-JUN-2003; 2003US-0476632P.  
XX 09-JUN-2003; 2003US-0476641P.  
XX 08-JUL-2003; 2003US-0485217P.  
XX 08-JUL-2003; 2003US-0485218P.  
XX 08-JUL-2003; 2003US-0485223P.  
XX 08-JUL-2003; 2003US-0485224P.  
XX 08-JUL-2003; 2003US-0485325P.  
XX 08-JUL-2003; 2003US-0485359P.  
XX 14-JUL-2003; 2003US-0486446P.

PR 14-JUL-2003; 2003US-0486480P.  
PR 15-JUL-2003; 2003US-0486891P.  
PR 15-JUL-2003; 2003US-0486896P.  
PR 08-AUG-2003; 2003US-0493341P.  
PR 08-AUG-2003; 2003US-0493370P.  
PR 08-AUG-2003; 2003US-0493373P.  
PR 08-AUG-2003; 2003US-0493577P.  
XX  
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX  
XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D,  
PI Halenbeck RF, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y,  
PI Wong JCP, Wu G, Zhang H, Zeng C;  
XX  
XX WPI; 2004-365511/34.  
XX  
XX N-PSDB; ADN98725.  
XX  
XX New nucleic acid molecules, useful in preparing a composition for  
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral  
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or  
PT ulcerative colitis.  
XX  
XX Claim 14; SEQ ID NO 1109; 532pp; English.  
XX  
XX The invention relates to a nucleic acid molecule comprising a  
CC polynucleotide sequence or its complement that encodes a polypeptide. The  
CC nucleic acid is useful in preparing a composition for treating or  
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,  
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
CC heart disease or ulcerative colitis. This sequence corresponds to a  
XX protein of the invention.  
XX  
XX Sequence 664 AA;  
SQ  
Query Match 18.3%; Score 407; DB 8; Length 664;  
Best Local Similarity 32.6%; Pred. No. 2,9e-23;  
Matches 115; Conservative 46; Mismatches 138; Indels 54; Gaps 16;  
QY 23 KYIICYEGCDKAYNRPISLLEQHLRTHSNDPRKCTVDDCDKAFKSHLETIVGHSK 82  
DB 298 KPYQC--KECDKVFSSKSYLERHRIHTEGKPKCY--CDKAFRDSHLAQHIVIHTR 353  
QY 83 KPFHSGVCGVNSRQLKRNHEITHT--KSFCTFENCOEAFYKHSL-RHHISVHEKT 139  
DB 354 KPYKNECGKTREBNALVHKTIHTEGKPKYC--NECGKVFQGSNLARHRIHTEGK 411  
QY 140 LTCKQCNKVFTRPSKLAQHLKHHGSPAYOCDPHCFFNPTQWVLAQFH--IKOSHPRK 197  
DB 412 YCKECCDKVFSRSHLERHRIHTEGKPKYC--KYCDKAFRDSHLAQHIVIHTEGK 468  
QY 198 KCPKCGKCGVCKGLSSHMLSHDDSTMIKIMTCIDYDVCK-PAKKNELVEHNIPIFDGNI 256  
DB 469 KNECGKTFVONSILVHMKVIH---TGEKRYKNEC--KSFPHKSLVYHRIH--HTGK 522  
QY 257 PDDLKETEYKLENLDDGSKLNNLEHETELKYEEDEDEDEDLDEKRSVDRSDMS 316  
DB 523 P---YKNECGKVFQGSQSLACHRIHTEGK-KPYKEEDC----- 559  
QY 317 AGRSIKFTASLEGSKSVSLISNGSKTINCPKNCNDRMFSREYDLRRHLKMH 369  
DB 560 -----VFNIKSHLEIHRVHTEGKPKYC--RVCDKAFGRDSYLAQHORVH 602  
RESULT 7  
AAU30831  
ID AAU30831 standard; protein; 1230 AA.  
XX  
XX AAU30831;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX Novel human secreted protein #1322.  
XX



KW Human; vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 OS Homo sapiens.  
 XX WO200179449-A2.  
 XX EN  
 XX PD 25-OCT-2001.  
 XX PF 16-APR-2001; 2001WO-US008656.  
 XX PR 18-APR-2000; 2000US-00552929.  
 XX PR 26-JAN-2001; 2001US-00770160.  
 XX PA (HYSB-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Dmanac RT;  
 XX DR WPI; 2001-611725/70.  
 XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 XX PT vaccination, testing and therapy.  
 XX PS Claim 20; Page 355; 765pp; English.  
 XX CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising the  
 CC nucleic acids encoding the polypeptides and cells genetically engineered  
 CC to express them are also useful for producing the proteins. The proteins  
 CC are useful in genetic vaccination, testing and therapy, and can be used  
 CC as nutritional supplements. They may be used to increase stem cell  
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
 CC and/or nerve tissue growth or regeneration; immune suppression and/or  
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
 CC AAU28510-AAU33304 represent the amino acid sequences of novel human  
 CC secreted proteins of the invention  
 XX  
 XX SQ Sequence 1230 AA;  
 Query Match 18.1%; Score 403; DB 4; Length 1230;  
 Best Local Similarity 32.9%; Pred. No. 1,4e-22;  
 Matches 118; Conservative 45; Mismatches 144; Indels 52; Gaps 17;  
 QY 20 SRPKYICTYEGCDKAYNRPSLLBOHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSH 79  
 DB 891 TREKPYKC--EEGKAFSPQPSHLTHKRNHTEKPYKC--EEGKAFSSSTLTTHKIIH 946  
 QY 80 SEKKPHGCVCGKGVNSRQHLKHEITHT--KSFKCTFENCQEAIFYKQSL-RHHILSVH 136  
 DB 947 TSEKPYKCEGCKAFKSSSTLTENKIIHTEGKPYKC--EEGKAFSSSTLTTRHTMHNG 1004  
 QY 137 EKTLTQCKNKVTPTRSKLAQHLKHHGSPAYQCHPGCFKPFQWVSVLQPHIKQSHPK 196  
 DB 1005 EKPYPKEEGCKAFNRSSKLTTHKIIHTEGKP-YKCE--CGKAFSSSTLTNGH-KKIHNR 1060  
 QY 197 LKCPKCGKGCVGK-----KGLSHMLSHDSTMIKIWTCDYCVGKFAKQNELVHNYI 250  
 DB 1061 EKPYPKC-BGC-GKAFSQSPFTLTGKRIH--TGKPYKCGEC--GKAKESALYKHI 1113  
 QY 251 FHDGNIIPDDLKETEYKLENIIDQSKLNLIHLETEKLKVEEDEDSDIDEKRSDV 310  
 DB 1114 IHTGE-----KPYKCECKCKAFNQSILTNHKKIHTIPKIH-----T 1151  
 QY 311 RSDNSAQHSIKSFTNLSGSKSVSKLISNGKINCPCPNCDRNFSSRYDLRRHLKWH 369  
 DB 1152 REKPYKYKCGKGSFNRSSTFTK--HKVITHGVGLYKC--EEGCKSFFWSSALTRHKIH 1206

RESULT 8  
 ADM05916  
 ID ADM05916 standard; protein; 404 AA.  
 XX  
 XX AC ADM05916;  
 XX AD 20-MAY-2004 (first entry)  
 XX DE Human protein of the invention SEQ ID NO:4601.  
 KW human; gene therapy; diagnostic marker; pharmaceutical.  
 OS Homo sapiens.  
 XX EP147046-A1.  
 XX PN 24-SEP-2003.  
 XX PD 12-APR-2002; 2002EP-00008400.  
 XX PF 22-MAR-2002; 2002JP-00137785.  
 XX PR (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX PA  
 XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Mauno Y;  
 DR WPI; 2003-723558/69.  
 DR N-Psdb; ADM03473.  
 XX PT New polynucleotides and polypeptides are useful in gene therapy, for  
 PT developing a diagnostic marker or medicines for regulating their  
 PT expression and activity, or as a target of gene therapy.  
 XX PS Claim 1; SEQ ID NO 4601; 305pp; English.  
 XX CC The invention relates to a novel human polynucleotide and the encoded  
 CC polypeptide. A polynucleotide of the invention may have a use in gene  
 CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful  
 CC as a primer for synthesizing the polynucleotide or as a probe for  
 CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are  
 CC useful in gene therapy, for developing a diagnostic marker or medicines  
 CC for regulating their expression and activity, or as a target of gene  
 CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides  
 CC are useful as pharmaceutical agents. The present sequence represents a  
 CC protein sequence of the invention.  
 XX  
 XX SQ Sequence 404 AA;  
 Query Match 17.9%; Score 400; DB 7; Length 404;  
 Best Local Similarity 33.9%; Pred. No. 5.2e-23;  
 Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;  
 QY 23 KVIITVYGCPCAYNRPSLLBOHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSHSEK 82  
 DB 39 KPYKC--EEGKAFNRSSILTHKIIHTEGKPYKC--EEGKGSSTLTTHKAIHABE 94  
 QY 83 KPHGCVCGKGVNSRQHLKHEITHT--KSFKCTFENCQEAIFYKQSLRHHILSVH--EK 138  
 DB 95 KPYKCEEGCKASNSSSKLMEHKRIHTEGKPYKC--EEGKAFSSSTLTEN-KRIHAGEK 151  
 QY 139 TLTQCKNKVTPTRSKLAQHLKHHGSPAYQCHPGCFKPFQWVSVLQPH--IKQSHPK 196  
 DB 152 PYKCEEGCKAFNRSSILTHKIIHTEGKP-YKCE--GGCKAFSKVSTLTNTKAIHABEKP 208  
 QY 197 LKCPKCGKGCVGKGLSHMLSHDSTMIKIWTCDYCVGK-FAKQNELVHNYI FHDGN 255  
 DB 209 YKCEGKASNSSSKLMEHKRIH--TGKPYKCEB--GKAFSSSTLTENKRI-HAGE 262  
 QY 266 IPDDLKETEYKLENIIDQSKLNLIHLETEKLKVEEDEDSDIDEKRSDVRSDS 314

Db 263 KP---YKCECGKAFWSSSFTGKRIHAE-KPYKCECGKGFSTFSLTGHKIHTGE 318  
 QY 315 --MSAQRSTKSTFASLEGSKSVKLSNGKINCPCNNDMPFSEYDLRHLKWH 369  
 Db 319 KPYKCECGKAF--SWSSILTEHKIHTGEKPYKC--EECGKAFSRSSSLTHKRIH 371

## RESULT 9

ADN06034  
 ID ADN06034 standard; protein; 803 AA.

AC ADN06034;

DT 01-UTL-2004 (first entry)

DE Antipsoriatic protein sequence #1173.

KW antipsoriatic; gene therapy; psoriasis; diagnosis.

OS Homo sapiens.

PN WO2004028479-A2.

PD 08-APR-2004.

PF 25-SEP-2003; 2003WO-US030907.

PR 25-SEP-2002; 2002US-0414006P.

PA (GETH) GENENTECH INC.

PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;

PI Wu TD;

DR WPI; 2004-305105/28.

DR N-PSDB; ADN06033.

PT New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.

PS Claim 9; SEQ ID NO 2429; 3069bp; English.

CC The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polypeptides of the invention.

CC Sequence 803 AA;

Query Match 17.9%; Score 400; DB 8; Length 803;

Best Local Similarity 31.9%; Pred. No. 1,3e-22; Indels 34; Gaps 19;

Matches 121; Conservative 53; Mismatches 171; Indels 34; Gaps 19;

QY 2 SESDETKSISLSSSS--SRPKYICTYEGCDKAYNPSLLBOHLRTHSNDPRYKCTV 59

Db 367 TEGGEAFSRSSNLTGKRIHTGEKPYKCECGKAFNPSLTHKRIHTGEKPYKC--EE 422

QY 60 DDCDKAFKRSKSHLETHIVSHSEKKPFHSCVCGKVSROGLKHEITHT--KSFCTFEN 117

Db 423 EECGKAFNWPSTLTGKRIHTGEKPYKCECGKAFNPSLTHKRIHTGEKPYKC--EE 480

QY 118 COEAFYFGHSL-RHILSVHKEKTLTKQCNKVFTRPEKLAQHKLKHGGSPAYQCDHPC 176

Db 481 CGKAFSSSNLTGKRIHTGEKPYKCECGKAFNPSLTHKRIHTGEKPYKC--EE 537

QY 177 FKQFQTSVLQFH--IKQSHPKLKCPKCGKGVGKGLSSHMLSHDSTWIKIWTCDYCD 234

Db 538 GKAFNHSILTKKRIHTGEKPYKCECGKAFNPSLTHKRIHTGEKPYKC--EE 593

QY 235 VGR-FAKKNELVHYNI-FHONGIPDDLKTEYVKLENLDDGSKANNLHLETKLAVE 293

Db 594 -GKAFQSSNLTGKRIHTGEKPYKCECGKAFNPSLTHKRIHTGEKPYKC 647  
 QY 294 E-DEEDSDSLDEKSDVRSDS--MSAQRSTKSTFASLEGSKSVKLSNGKINCPCN 350  
 Db 648 ECGKAFNWPSTLTGKRIHTGEKPYKCECGKAF--KLSTLTGKRIHTGEKPYKC 704  
 QY 351 NCDPMFSREYDLRHLKWH 369  
 Db 705 -CGKAFNPSNLTGKRIHTGEKPYKCECGKAF--KLSTLTGKRIHTGEKPYKC 722

## RESULT 10

AA038689  
 ID AA038689 standard; protein; 809 AA.

AC AA038689;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 1834.

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-UTL-2001.

PF 26-DEC-2000; 2000WO-US034263.

PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

PA (HYSR-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;

PI Zhou P, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.

DR N-PSDB; AA157845.

PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.

PS Example 3; SEQ ID NO 1834; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AA138642-AA42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,



CC preventing inflammatory, CNS, immune, bacterial or viral disorder,  
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
CC heart disease or ulcerative colitis. This sequence corresponds to a  
CC protein of the invention.

XX  
SQ Sequence 444 AA;

Query Match 17.9%; Score 399; DB 8; Length 444;

Best Local Similarity 30.8%; Pred. No. 7, 1e-23;

Matches 120; Conservative 47; Mismatches 179; Indels 44; Gaps 18;

QY 20 SRPKYICTYEGCDKAYNRPSLLLEOHILRTSHNDPRPKYCTVDDCKAFPRKSHLETIVSH 79  
DB 77 SGEKPYKC--KEGKAVNEASNLSTHKRIHTGKKPYKC--EEGKAFNRSLHTTKIILH 132  
QY 80 SEKKPFPGCVGKGVNSRQHLKHEITHT--KSPFKCTFENCQEAFFYHGL--RHHILSVH 136  
DB 133 TGGKPYKCEGCGKAFNOSANILTHHKRIHTGKPKYC--EEGKAFNOSSTLTAKIILHAG 190  
QY 137 EKLTLTGCKMKNVFRPSKLAQHKLKHGGSFAYOCDHPGCFKNGQTWSVLQPH--IKQSH 194  
DB 191 EKPYKCEGCGKAFNOSSTLTTHKI--HTGKPYKCEB--CGKAFSRSLHTTKRIHSGE 247  
QY 195 PKLKCPKCGKGVGKGLSSHMLSHDDSTWIKIWTCDYCDVGKFAKNEIYEHYNIFFHDG 254  
DB 248 KPYKCEGCGKAFNOSSTLTTHKRIHAG--KPYKCEVCSKA--FSRSHLTTHKRI--HTG 302  
QY 255 NIPDDLKETBYKLENLDDGSKLNNLHELETKEKLV--DEDEEDSLDEKSDVRSD 313  
DB 303 EKP---YKCEGCGKAFNLSQTLTHKRIHTGE--KPYKCEGCGKAFNOSSTLSKHKVIHTG 358  
QY 314 S--MSAQRSLKSFPAISLEGSKSVSKLISNGSKINCNCDDHFSREYDLRHHLMKWD-- 370  
DB 359 EKPYKCEGCGKAFNOS--SHLTTHKMIHTGKPKYC--EEGKAFNNSITLNRHKMIHTG 414  
QY 371 -----DNLQRISEFLNSIEKEE 387  
DB 415 EKLKPEPCNNACNMIAKISKYKNCNGEK 444

RESULT 12  
ADN99690  
ID ADN99690 standard; protein: 444 AA.  
XX  
AC ADN99690;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Novel human protein sequence #506.  
XX  
KW anti-inflammatory; dermatological; neuroprotective; immunomodulator;  
KW antibacterial; virucide; antipsoriatic; cytosolic; gene therapy;  
KW vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;  
KW early aging; hormonal imbalance; ischemic heart disease;  
KW ulcerative colitis.  
XX  
OS Homo sapiens.  
XX  
PN WO2004038003-A2.  
XX  
PD 06-MAY-2004.  
XX  
PF 24-OCT-2003; 2003WO-US033947.  
XX  
PR 25-OCT-2002; 2002US-0421061P.  
PR 25-OCT-2002; 2002US-0421080P.  
PR 25-OCT-2002; 2002US-0421552P.  
PR 25-OCT-2002; 2002US-0421614P.  
PR 30-OCT-2002; 2002US-0422177P.  
PR 30-OCT-2002; 2002US-0422178P.  
PR 30-OCT-2002; 2002US-0422179P.  
PR 15-NOV-2002; 2002US-0426355P.  
PR 15-NOV-2002; 2002US-0426384P.  
PR 15-NOV-2002; 2002US-0426394P.

PR 15-NOV-2002; 2002US-0426430P.  
PR 15-NOV-2002; 2002US-0426916P.  
PR 27-NOV-2002; 2002US-0429224P.  
PR 27-NOV-2002; 2002US-0429275P.  
PR 27-NOV-2002; 2002US-0429302P.  
PR 27-NOV-2002; 2002US-0429326P.  
PR 27-NOV-2002; 2002US-0429651P.  
PR 04-DEC-2002; 2002US-0430645P.  
PR 04-DEC-2002; 2002US-0430651P.  
PR 04-DEC-2002; 2002US-0430657P.  
PR 04-DEC-2002; 2002US-0430663P.  
PR 04-DEC-2002; 2002US-0430688P.  
PR 04-DEC-2002; 2002US-0430684P.  
PR 05-DEC-2002; 2002US-0430937P.  
PR 05-DEC-2002; 2002US-0430965P.  
PR 05-DEC-2002; 2002US-0431458P.  
PR 12-DEC-2002; 2002US-0433251P.  
PR 12-DEC-2002; 2002US-0433500P.  
PR 13-DEC-2002; 2002US-0433318P.  
PR 13-DEC-2002; 2002US-0433318P.  
PR 23-DEC-2002; 2002US-0436236P.  
PR 03-JAN-2003; 2003US-0437914P.  
PR 17-JAN-2003; 2003US-0440820P.  
PR 18-APR-2003; 2003US-0463700P.  
PR 18-APR-2003; 2003US-0463708P.  
PR 18-APR-2003; 2003US-0463716P.  
PR 18-APR-2003; 2003US-0463732P.  
PR 02-MAY-2003; 2003US-0467199P.  
PR 02-MAY-2003; 2003US-0467201P.  
PR 02-MAY-2003; 2003US-0467203P.  
PR 02-MAY-2003; 2003US-0467230P.  
PR 19-MAY-2003; 2003US-0471306P.  
PR 19-MAY-2003; 2003US-0471336P.  
PR 22-MAY-2003; 2003US-0472420P.  
PR 09-JUN-2003; 2003US-0476609P.  
PR 09-JUN-2003; 2003US-0476621P.  
PR 09-JUN-2003; 2003US-0476632P.  
PR 09-JUN-2003; 2003US-0476641P.  
PR 08-JUL-2003; 2003US-0485217P.  
PR 08-JUL-2003; 2003US-0485218P.  
PR 08-JUL-2003; 2003US-0485223P.  
PR 08-JUL-2003; 2003US-0485224P.  
PR 08-JUL-2003; 2003US-0485325P.  
PR 08-JUL-2003; 2003US-0485359P.  
PR 14-JUL-2003; 2003US-0486446P.  
PR 14-JUL-2003; 2003US-0486480P.  
PR 15-JUL-2003; 2003US-0486891P.  
PR 15-JUL-2003; 2003US-0486960P.  
PR 08-AUG-2003; 2003US-0493341P.  
PR 08-AUG-2003; 2003US-0493370P.  
PR 08-AUG-2003; 2003US-0493573P.  
PR 08-AUG-2003; 2003US-0493577P.

(FIVE-) FIVE PRIME THERAPEUTICS INC.

XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D,  
PI Halenbeck RF, Kotchakota S, Lin H, Linemann T, Pierce K, Wang Y,  
PI Wong JGP, Wu G, Zhang H, Zeng C;

DR WPI; 2004-365511/34.  
DR N-PSDB; ADN98906.

XX New nucleic acid molecules, useful in preparing a composition for  
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral  
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or  
PT ulcerative colitis.

PS Claim 14; SEQ ID NO 1290; 532pp; English.

XX The invention relates to a nucleic acid molecule comprising a  
CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or  
CC preventing inflammatory, CNS, immune, bacterial or viral disorder.  
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
CC heart disease or ulcerative colitis. This sequence corresponds to a  
CC protein of the invention.

XX Sequence 444 AA:

Query Match 17.9%; Score 398; DB 8; Length 444;  
Best Local Similarity 30.2%; Pred. No. 8.5e-23;  
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPKYICTYEGCDKAYNRPSSLEQHLRTSNDRPYKCTVDDCDKAFPRKSHLETHIVSH 79  
DB 77 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKPYKC--EECGKAFNRLSHLTTHKRIIH 132  
QY 80 SEKKPFHSGVCGKGNVSRHLLKHEHTHT--KSPKCTENCQEAFFKQSL--RHHLISLVH 136  
DB 133 TGKKPYKCECGKAFNOSANLTTTHKRIHTGKPYKC--EECGKAFNOSSTLTTHKRIIHAG 190  
QY 137 EKTITGKOCNKVTPRSKLAOHKLKHGSPAYQCDHPGCFKPNFTWVLAQFH--IKOSH 194  
DB 191 EKRYKCEBGKAFNOSSTLTTHKRI--HTGKPYKCE--CGKAFSOLSHLTTHKRIHSGE 247  
QY 195 PKLKCPKCGKCGVGGKGLSHMLSHDSTMIKIWTCDYCDVGKPAKNELVEHYNI PHDG 254  
DB 248 KPYKCECGKAFNOSSTLTTHKRIHAGE--KPYKCEVCSKA--FSRFSHLTHKRI--HTG 302  
QY 255 NIPDDLKETEYKLENLIDGSKLNNHLETEKLVKVEDDEDESDLSDEKSDVRSDS 314  
DB 303 EKP---YKCEBGKAFNLSQLTTHKRIHTGE--KPYKCEC-----GKAFNOSST 348  
QY 315 MSAORSI-----KSFASLEGSKSVSKLINSKKINCPKNNCDRMPFSREYD 361  
DB 349 LSKHKVITHGEKPYKYKECGKAFNOS--SHLTTHKMIHTGKPYKC--EECGKAFNNSI 404  
QY 362 LRRHLKMD-----DNLQRIEFLNSIEKEE 387  
DB 405 LNRHKMIHTGEKLYKDESCNNAADNIATISKYKRNAGAEK 444

RESULT 13

ID ABO60215 standard; protein; 487 AA.

XX ABO60215;

DT 29-JUL-2004 (first entry)

XX Human genome derived single exon protein #6449.

DE Human; gene expression; single exon probe; microarray;

KM alternative splicing event; genomic alteration.

XX Homo sapiens.

OS US2003194704-A1.

XX 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.

PR 03-APR-2002; 2002US-00029386.

PA (PENN/) PENN S G.

PA (HANK/) HANK D R.

PI Penn SG, Rank DR, Hanzel DK,

DR WPI, 2004-119264/12.

PT New human genome-derived single exon nucleic acid probes useful for human

PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.

PS Claim 45; SEQ ID NO 33849; 80pp; English.

XX The invention relates to a nucleic acid probe for measuring human gene  
XX expression, comprising any of the 27,400 fully defined nucleotide  
XX sequences in the specification, or their complements or fragments, and  
XX encoding at least 8 amino acids of any of the 688 amino acid sequences  
XX fully defined in the specification. The probe is a single exon probe that  
XX hybridizes under high stringency conditions to a nucleic acid molecule  
XX expressed in human cells or tissues. Also included are a spatially-  
XX addressable set of single exon nucleic acid probes for measuring human  
XX gene expression (comprising a plurality of single exon nucleic acid  
XX probes cited above, where each of the plurality of probes is separately  
XX and addressably isolatable or amplifiable from the plurality), a single  
XX exon microarray for measuring human gene expression, a method of  
XX measuring human gene expression, a vector comprising the single exon  
XX probe cited above, an ORF-encoded peptide comprising at least 8  
XX contiguous amino acids of any of the above-mentioned amino acid  
XX sequences (optionally with conservative amino acid substitutions), an  
XX isolated antibody that binds specifically to a peptide cited above,  
XX methods of selling and/or licensing single exon probes or microarrays to  
XX a customer desiring to measure gene expression, a method of providing  
XX human gene expression data by subcription, and a computer-readable  
XX storage medium which contains a database having a plurality of records  
XX (each record including data on the expression of a single exon probe  
XX cited above. The probe, method and apparatus are useful in gene  
XX expression analysis. The probes may be used as tools for surveying  
XX tissues to detect the presence of expressed messages that contain their  
XX specific exon, or in constructing genome-derived single exon microarrays.  
XX In addition, the probes are used in identifying and characterizing  
XX alternative splicing events, in detecting and characterizing gross  
XX alterations in the genomic locus that includes their exon, in assessing  
XX smaller genomic alterations, in priming the synthesis of nucleic acids,  
XX or in expressing the ORF-encoded peptide. The present sequence is a human  
XX single exon probe protein of the invention. Note: The sequence data for  
XX this patent did not form part of the printed specification, but was  
XX obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?docID=20030194704

XX Sequence 487 AA;

Query Match 17.9%; Score 398; DB 8; Length 487;  
Best Local Similarity 30.2%; Pred. No. 9.7e-23;  
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPKYICTYEGCDKAYNRPSSLEQHLRTSNDRPYKCTVDDCDKAFPRKSHLETHIVSH 79  
DB 120 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKPYKC--EECGKAFNRLSHLTTHKRIIH 175  
QY 80 SEKKPFHSGVCGKGNVSRHLLKHEHTHT--KSPKCTENCQEAFFKQSL--RHHLISLVH 136  
DB 176 TGKKPYKCEBGKAFNOSANLTTTHKRIHTGKPYKC--EECGKAFNRLSHLTTHKRIIHAG 233  
QY 137 EKTITGKOCNKVTPRSKLAOHKLKHGSPAYQCDHPGCFKPNFTWVLAQFH--IKOSH 194  
DB 234 EKRYKCEBGKAFNOSSTLTTHKRI--HTGKPYKCE--CGKAFSOLSHLTTHKRIHSGE 290  
QY 195 PKLKCPKCGKCGVGGKGLSHMLSHDSTMIKIWTCDYCDVGKPAKNELVEHYNI PHDG 254  
DB 291 KPYKCEBGKAFNOSSTLTTHKRIHAGE--KPYKCEVCSKA--FSRFSHLTHKRI--HTG 345  
QY 255 NIPDDLKETEYKLENLIDGSKLNNHLETEKLVKVEDDEDESDLSDEKSDVRSDS 314  
DB 346 EKP---YKCEBGKAFNLSQLTTHKRIHTGE--KPYKCEC-----GKAFNOSST 391  
QY 315 MSAORSI-----KSFASLEGSKSVSKLINSKKINCPKNNCDRMPFSREYD 361  
DB 392 LSKHKVITHGEKPYKYKECGKAFNOS--SHLTTHKMIHTGKPYKC--EECGKAFNNSI 447  
QY 362 LRRHLKMD-----DNLQRIEFLNSIEKEE 387

Db 448 LNRHMKIHTGEKLYKPESCNNACDNIATKISKYKRNCAGEK 487

RESULT 14

ABU96690 standard; protein; 563 AA.

ABU96690;

25-JUL-2003 (first entry)

Human nucleic acid-associated protein (NAAP) #19.

Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic; anticonvulsant; neurotropic; neuroprotective; cerebroprotective; anti-HIV; anti-allergic; anti-inflammatory; chymotrypsin; gene therapy; cell proliferative disorder; cancer; atherosclerosis; neurological disorder; epilepsy; Huntington's disease; stroke; immune disorder; inflammatory disorder; AIDS; allergy; developmental disorder; Hypochyroidism; Cushing's syndrome; infection; protein-protein interaction; drug-target interaction; gene expression profile.

Homo sapiens.

MO2003023003-A2.

20-MAR-2003.

05-SEP-2002; 2002MO-US028540.

07-SEP-2001; 2001US-0317792P.  
07-SEP-2001; 2001US-03177912P.  
14-SEP-2001; 2001US-0322270P.  
21-SEP-2001; 2001US-0324040P.  
28-SEP-2001; 2001US-0326732P.  
19-OCT-2001; 2001US-0346716P.  
25-JAN-2002; 2002US-0351749P.  
22-FEB-2002; 2002US-0359498P.

(INCY-) INCYTE GENOMICS INC.

Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe LJ, Becha SD, Richardson TW, Lee EA, Sprague WR, Emerling BW, Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Li XJ, Hafalia AJA, Sanjanwala B, Marquis JP, Gorvay AE, Lee SY, Ison CH, Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebardjian Y, Shah P, Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE, Burford N, Rankumar J;

WPI; 2003-313243/30.  
N-PSDB; ACA98938.

New human nucleic acid associated proteins (NAAP), useful for diagnosing, treating and preventing diseases or conditions associated with the aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or infections.

Claim 1; Page 255-257; 345pp; English.

The invention describes a novel human isolated nucleic acid-associated polypeptide (NAAP). The polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of NAAP, such as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome) disorders, or infections. These are also useful in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of NAAP. The NAAP or its fragments are useful in screening compounds for effectiveness as agonist or antagonist of the polypeptides, or in altering the expression of the target polynucleotide and compounds

CC that specifically bind to or modulate the activity of the polypeptide.  
CC The microarray is useful in monitoring or measuring protein-protein  
CC interactions, drug-target interactions, and gene expression profiles.  
CC This is the amino acid sequence of a novel human nucleic acid-associated  
CC protein (NAAP)  
XX  
XX  
SQ Sequence 563 AA;

Query Match 17.9%; Score 398; DB 6; Length 563;  
Best Local Similarity 30.2%; Pred. No. 1,2e-22;  
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SPPKTYICTEGCDKAYNRPSSLECHLRTHSNDPRKYCVDDCKAFPKSHLETHIVSH 79  
Db 196 SGEKPYKC--KECGKAYNTSNLSHTKRIHTGKKPYKC--EECGKAFNNLSHTTKRIH 251  
QY 80 SEKKPFGVCGGVNSRQHLKRHEITHT--KSFCTFENCGAPYKQSL-RHILSYH 136  
Db 252 TGKKPYKCECGKAFNQSANTLTHKRIHGEKPYKC--EECGKAFQSSTLTFAKRIHAG 309  
QY 137 EKLTCCKCNKYPTRPSKLAQHKLNHGGSPAVQCDHPGCFNPFQWVSLQFH--IKQSH 194  
Db 310 EKPYPKCECGKAFQSSTLTTHKI-IHTGEKPYKCE--CGKAFSQLSHLTTHKRIHSGE 366  
QY 195 PKLKCPKCGKCGVKKGLSSHMLSHDSTMIKIWTCDYCVGKFAKNELVEHYNIFFHQ 254  
Db 367 KPYKCECGKAFKQSSLTTHKRIHAG--KPYKCEVSKA-FSPSHLTTHKRI-HIG 421  
QY 255 NIPDDLKTEYVKLENLLDQSKLNINLEHLETEKLVEDBEDBEDSDVRS 314  
Db 422 EKP--YKCECGKAFNLSQLTTHKIHTGE-KPYKCEC-----GKAFQSSST 467  
QY 315 MSAQSI-----KSFASLEGSKSVSKLINSKKINCPKNCNDRMFSREYD 361  
Db 468 LSKHVIHTGEKPYKCECGKAFNOS--SHLTTHKRIHGEKPYKC--EECGKAFNNSI 523  
QY 362 LRRHLKMH-----DNLQRIESFNSIEKEE 387  
Db 524 LNRHMKIHTGEKLYKPESCNNACDNIATKISKYKRNCAGEK 563

RESULT 15

ADI60246 standard; protein; 563 AA.

ADI60246;

15-APR-2004 (first entry)

Secreted polypeptide #130.

osteopathic; vulnery; cytosolic; gene therapy; diagnosis; forensics;  
KW gene mapping; mutation identification; biodiversity; chromosome marker;  
KW immune response; myeloid cell disorder; lymphoid cell disorder;  
KW bone cartilage; tendon; ligament; nerve tissue growth; wound healing;  
KW burns; incision; ulcer; cancer.

Homo sapiens.

MO2003025142-A2.

27-MAR-2003.

18-SEP-2002; 2002MO-US029636.

18-SEP-2001; 2001US-0323349P.

16-SEP-2002; 2002US-00323349.

(HYSE-) HYSEQ INC.

Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J, Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Dmanac RT;

DR WPI; 2003-354601/33.  
DR N-PSDB; ADI60591.

DR N-PSDB; ADI60591.

PT New polynucleotides and secreted proteins, useful for treating myeloid or lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve tissue growth or regeneration, in wound healing, and in tissue repair and/or replacement.

PI replacement.

PI replacement.

PI replacement.

PS Claim 20; SEQ ID NO 281; 243pp; English.

The invention relates to novel isolated polynucleotides or a sequence encoding a polypeptide with biological activity, where the polynucleotide hybridizes to the polynucleotide under stringent hybridization conditions or has greater than 99% sequence identity with the polynucleotide. The polynucleotides and polypeptides are useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders and other traits, to assess biodiversity, as nutritional sources or supplements. The polynucleotides may also be used as molecular weight markers, chromosome markers or map related gene positions, or as an antigen to raise anti-DNA antibodies or elicit immune response. The polypeptides are useful for raising antibodies, as markers for tissues in which the corresponding polypeptide is expressed, for re-engineering damaged or diseased tissues, for treating weldoid or lymphoid cell disorders, in bone cartilage, tendon, ligament and/or nerve tissue growth or regeneration, in wound healing, in tissue repair and replacement, in healing of burns, incisions and ulcers, and in treating cancer. This sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

CC sequence corresponds to a protein sequence of the invention.

**SQ** Sequence 563 AA;

Query Match 17.9%; Score 398; DB 7; Length 563;

Query Match 17.9%; Score 398; DB 7; Length 563;

Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPQKVCITBEGCDKAYNPSLLEOHAFHNSDRKYCTVDDCPKAPRKSHEITHVSH 79

Db 196 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTHKRIH 251

QY 80 SEKKPFHCVSGVNSRHLKRHEIYTH--KSFCTEPCOZAFYKQSL--RHHLISVH 136

Db 252 TGKKPYKCECGKAFNQSANLTHHKRIHTEKPYKC--EECGKAFNQSSTLTANKRIHAG 309

QY 137 EKLTVTCQCNKVITRPSKLAQHLKHHGSPAYOCDFHCFTRNQTSVLQFH--IKQSH 194

Db 310 EKEPYKECECGKAFNQSSTLTTHKRI--IHTGEKPYKCEB--CGKAFNQLSHLTHHKRIHSGE 366

QY 195 PKLCPKPCGCGCAGKGGKSSHLMSHDSDSYIMLIMTCDDVCGKFAKXNLEVENHIEHDG 254

Db 367 KPYKCECGKAFNQSSTLTTHHKRIHAGE--KPYKCEVCSRA--FSRSHLTHHKRI--HTG 421

QY 255 NIPDDLKEREVEKLLNLLDQSKLNNLHELTEKLYKEEDBEDBEDSLDEKRSADVSDS 314

Db 422 EKP---YKCECGKAFNLSQLTTHKRIHTEG--KPYKCEB-----GKAFNQSST 467

QY 315 MSARST-----KSFYASLEGSKSVSLKINSKKINCPCPNDCRMFSREYD 361

Db 468 LSKHKVHITGEKPYKCECGKAFNQS--SHLTHHKRIHTEKPYKC--EECGKAFNNSSI 523

QY 362 LRRLKMHND-----DNLQRIESTLNSIEKEE 387

Db 524 LNRHKMITHGEKLYKPECSNNACNNIAKISYKRNCAAGEK 563

Db 196 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIH 251

80 SEKKPFHCVCCKGVNSRQHLKRHEITHT--KSFKCTFENCQEA FYKHQSL-RHILSVH 136

Db 252 TGKKPYKCECGKAFNQSANLTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKI IHAG 309

QY 137 EKTLCQCNKVFTRPSKLAQHKLKHGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194

Db 310 EKP YKCEECGKAFS Q S S T L T H K I - I H T G E K F Y K C E E - - C G K A F S Q L S H L T T H K R I H S G E 366

QY 195 PKLKPCGKGCVKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254

Db 367 KPYKCECGKAFKQSTLTTHKRIHAGE--KPYKCEVCSKA-FSRFSLTTHKRI-HTG 421

QY 255 NIPDDLKETEVKKLENLDQGSKLNNLHELJTEKLKVEEDEDEEDSLDEKRSVRSDS 314

Db 422 EXP---YKCECGKAFNLSSQLTHKIHTGE-KPYKCEC-----GKAFNQST 467

QY 315 MSAQRSI-----KSF TASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361

Db 468 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 523

QY 362 LRRHLKWD-----DNLQRIESFLNSIEKEE 387

Db 524 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 563

Search completed: February 8, 2005, 15:18:59  
Job time : 79 secs

Job time : 79 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:14:56 ; Search time 42 Seconds  
(without alignments)  
732.272 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229  
Sequence: 1 MSSEDTKSISSLISSSSSS.....PLVKARMDLPNETSVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCtus\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1036	46.5	208	4 US-09-248-796A-19287	Sequence 19287, A
2	671	30.1	153	4 US-09-248-796A-19289	Sequence 19289, A
3	400	17.9	803	4 US-09-538-092-1026	Sequence 1026, Ap
4	399.5	17.9	656	4 US-09-949-016-10075	Sequence 10075, A
5	395.5	17.7	423	1 US-08-523-376-3	Sequence 3, Appl1
6	393.5	17.7	1191	4 US-09-949-016-6356	Sequence 6356, Ap
7	358	16.1	647	4 US-09-389-956-68	Sequence 6, Appl1
8	358	16.1	647	4 US-09-389-956-6	Sequence 6, Appl1
9	351	15.7	487	4 US-09-949-016-10055	Sequence 10055, A
10	350	15.7	685	4 US-09-949-016-8627	Sequence 8627, Ap
11	344	15.4	639	4 US-09-949-016-9434	Sequence 9434, Ap
12	343	15.4	746	4 US-09-949-016-11494	Sequence 11494, A
13	343	15.4	769	4 US-09-949-016-8149	Sequence 8149, Ap
14	343	15.4	810	4 US-09-538-092-11275	Sequence 11275, Ap
15	337	15.1	469	4 US-09-538-092-1195	Sequence 1195, Ap
16	333	14.9	642	4 US-09-949-016-6986	Sequence 6986, Ap
17	332	14.9	589	4 US-09-949-016-10374	Sequence 10374, A
18	329	14.8	728	2 US-08-475-844-5	Sequence 5, Appl1
19	329	14.8	728	2 US-08-475-844-5	Sequence 5, Appl1
20	329	14.8	728	2 US-08-475-844-5	Sequence 5, Appl1
21	322	14.4	727	5 US-08-475-844-9	Sequence 9, Appl1
22	322	14.4	727	5 PCT-US95-08429-9	Sequence 9, Appl1
23	322	14.4	736	4 US-09-949-016-9241	Sequence 9241, Ap
24	321	14.4	654	4 US-09-538-092-1193	Sequence 1193, Ap
25	321	14.4	654	4 US-09-949-016-6357	Sequence 6357, Ap
26	321	14.4	663	4 US-09-949-016-7484	Sequence 7484, Ap
27	319.5	14.3	711	2 US-08-820-170A-10	Sequence 10, Appl

28	319.5	14.3	711	3 US-09-055-699-10	Sequence 10, Appl
29	319.5	14.3	711	3 US-09-273-565-10	Sequence 10, Appl
30	319.5	14.3	711	3 US-09-565-538-10	Sequence 10, Appl
31	319.5	14.3	711	3 US-09-661-468-10	Sequence 10, Appl
32	319.5	14.3	711	4 US-09-976-165-10	Sequence 10, Appl
33	319.5	14.3	711	4 US-09-828-648-2	Sequence 1356, Ap
34	317	14.2	321	4 US-09-538-092-1136	Sequence 1304, Ap
35	316.5	14.2	321	4 US-09-949-016-6364	Sequence 6364, Ap
36	316.5	14.2	321	4 US-08-933-750C-17	Sequence 17, Appl
37	316	14.2	488	3 US-09-234-613-17	Sequence 17, Appl
38	316	14.2	488	3 US-09-881-578A-4	Sequence 4, Appl1
39	316	14.2	314	4 US-09-949-016-8516	Sequence 8516, Ap
40	315.5	14.1	803	3 US-09-063-035-2	Sequence 2, Appl1
41	315	14.1	532	4 US-09-389-956-80	Sequence 80, Appl
42	314.5	14.1	506	4 US-09-538-092-1177	Sequence 1177, Ap
43	313	14.0	506	4 US-09-881-578A-2	Sequence 2, Appl1
44	313	14.0	717	4 US-09-362-123A-4	Sequence 4, Appl1
45	312.5	14.0	543	4 US-09-362-123A-4	Sequence 4, Appl1

ALIGNMENTS

```
RESULT 1
US-09-248-796A-19287
; Sequence 19287, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074, 725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096, 409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19287
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19287

Query Match          46.5%; Score 1036; DB 4; Length 208;
Best Local Similarity 96.2%; Pred. No. 3.8e-90;
Matches 202; Conservative 2; Mismatches 2; Indels 4; Gaps 2;

QY      204 KG-CVKKKGLSSHTLSDHDSMTKIKWTCDYCDYGRKAKKNEIYEHYNIFFDGNIPDDLK 262
          2 KGICLKKKKGLSSHTLSDHDTMKIKWTCDYCDYGRKAKKNEIYEHYNIFFDGNIPDDLK 61
          263 ETEVKKLENNLDGSKLNNLHELETEKLVKEDEDEDESDLSDEKRSDVDSNSAORSIK 322
          62 ETEVKKLENNLDGSKLNNLHELETEKLVKEDEDEDESDLSDEKRSDVDSNSAORSIK 118
QY      323 SFTASLEGSKSVKLSNSGKKINCPKNCNDRMFSEYDRLRRLLKMHDDNLQRIEFPNLS 382
          119 SFTASLEGSKSVKLSNSGKKINCPKNCNDRMFSEYDRLRRLLKMHDDNLQRIEFPNLS 178
QY      383 IEKEETPEGEPPLVKARMDLPNETSVISR 412
          179 IEKEETPEGEPPLVKARMDLPNETSVISR 208

RESULT 2
US-09-248-796A-19289
; Sequence 19289, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
```



```

1  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
2
3  FILE REFERENCE:  107196.112
4
5  CURRENT APPLICATION NUMBER:  US/09/248,796A
6
7  CURRENT FILING DATE:  1999-02-12
8
9  PRIOR APPLICATION NUMBER:  US 60/074,725
10
11 PRIOR FILING DATE:  1998-02-13
12
13 PRIOR APPLICATION NUMBER:  US 60/096,409
14
15 PRIOR FILING DATE:  1998-08-13
16
17 NUMBER OF SEQ ID NOS:  28208
18
19 SEQ ID NO 19289
20
21 LENGTH:  153
22
23 TYPE:  PRT
24
25 ORGANISM:  Candida albicans
26
27 US-09-248-796A-19289

```

Query Match	30.1%	Score 671;	DB 4;	Length 153;
Best Local Similarity	98.4%;	Pred. No. 9.9e-56;		
Matches 122;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	1	MSSDDEFKTSSLSISSSSSRPKXICTYEGCDKXVNPSSLLEOHLRHNSUDRKYCTVD	60
Db	18	MSSDDEFKTSSLSISSSSSRPKXICTYEGCDKXVNPSSLLEOHLRHNSUDRKYCTVE	77
Qy	61	DCDKAPFRKSHLETHVSHSEKKPHFCVSCGKGVNSRQHLKRHEITHTSKFCTFENCQE	120
Db	78	DCDKAPFRKSHLETHVSHSEKKPHFCVSCGKGVNSRQHLKRHEITHTSKFCTFENCQE	137
Qy	121	AFYK	124
Db	138	TFYK	141

```

RESULT 3
US-09-538-092-1026
; Sequence 1026, Application US/09538092
; Patent No. 675314
; GENERAL INFORMATION:
; APPLICANT: Glac, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurataseqFormatter Version 0.9
; SEQ ID NO 1026
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P28160
; US-09-538-092-1026

```

Query Match	17.9%;	Score 400;	DB 4;	Length 803;
Best Local Similarity	31.9%;	Pred. No. 5.4e-29;		
Matches 121;	Conservative 53;	Mismatches 171;	Indels 34;	Gaps 19;

```

Qy 2 SEEDETKSISSLSISSSS--SRPKYICYEECDKAYNRBLLLEOHLTSSNDAPRYCIV 59
Db 367 TEGGEAFSRSSNLTTHKRIKHTERKPKYC--EECGAFPMSSKLTENHKLTHGEKPYKC-- 422
Qy 60 DDCDKAFPRSHLETHIVSHSEKPPHSCVCGKGNVSRQHKRHEITH--KSPKCFEN 117
Db 423 EEEGKAFFMNPSTLYTKNRIHTGEKPKYCEVCGKAPQNSNLTTHKRIHTAEKPKYC--EE 480
Qy 118 CQEAFFKHQSL-RHHILSVHEKTLTCKQCNVFTPRSKLAQHKLKHGGSPAYQCDHPG 176

```

Db	441	CGKAFSRBSNLTKHKKIHIEKKPYCECGGKAFKMSKLTENKIHITGKPP--YKCE--C	537
QY	177	FKNFQTSVLQPH--IKOSHPLKCKPKCGKGVGKKGLSSHMLSHDSDTMIRKIWTCDYCD	234
Db	538	GKAFHFHFIILTHGKKIHITGEKPKYKCECGKAFQTSNLTHKKIH---TGKFPYKCEEC-	593
QY	235	VGK--FAKKNEIVEHNIPIHDGNIIPDDLKPEVTKLEMLDDGSKLNNIHELETKLKYE	293
Db	594	GKAFQTSNLTHKKI-HTGKPP--YKCECGAFNQSFTLTGKKIH-TEBEKPKCE	647
QY	294	B-DEEDEDSDLDKRRSDVRSDS--MSAQRISIKFASLSEGSKVSKLISNSGKINCPCPN	350
Db	648	EGKAFKMSSTLTGKKIHITGEKPKYKCECGAF--KLSSLTSTHKIHITGKPKYKCEK-	704
QY	351	NCDDRFSEFYDLRRLLKYN	369
Db	705	CGKAFNRPSNLTENKIH	722

```

RESULT 4
US-09-949-016-10075
; Sequence 10075, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10075
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10075

```

Query Match	17.9%	Score 399.5;	DB 4;	Length 656;
Best Local Similarity	29.7%	Pred. No. 4.5e-29;		
Matches 116;	Conservative 56;	Mismatches 158;	Indels 61;	Gaps 19;

```

OY 16 SSSSRPRPK-----YCTEGCDKANRPSLLQHLRTSNDNRPYCTVDDCQKAPFR 68
Db 57 SSSLTRHRKRIHTEGKPYKC--EECGKAFSHSSSTLAKHRKIHTEGKPYKC--EECGKAFR 112
OY 69 KSHLETHIVSHSEKKPFHSCVOCGKVNRSROHLKRHEITHT--KSEPKTCFENCOAFAFYKQ 126
Db 113 SSLAKHRKRIHTEGKPYKCCEGGAFAFSSSTLANHKIITHTEKPYKC--KEEDKTFKRLS 170
OY 127 SL-RHHILSVHEKTLTLCQCNKVPTRPSKLAQHLKHHGSGSPAYOCDHGCEPKNEQWTSV 185
Db 171 TLVTKHKIITHAEKLYKCEEGCAFNRSSNLTLHKRIHTEGKR--YKCEE--CGKAFNRSSS 227
OY 186 LQFHIKOSHPR--LKCPKCGKGCYGGKGLSSHMLSHDSTIMLKITWTDYCDVGR-PAVK 241
Db 228 LTKH-KRIHTREREKPFKCECGKAFWSSSTLTPRKRH---TGEKPYKCEEC--GRAFSRS 281
OY 242 NELVEHNIIFHDGNI P-----DDLKETRYKLENNLLDQSKL-----NNLHFL 285
Db 282 STLTKHKTl-ETGEKPYKCECGKAFKHSALAKRKIITHAEKLYKCEEGCAFQOQSSNL 340
OY 286 ETEKLYKVEDEDEDESLDE-----KRSVDPSMSAQRISFTASLEGSXSV 334
Db 341 TTKHKIITHKEKPSSEEDCKAFIWSSTLTENKRIITHREKTYKCEEGCAF--SQPSHLTT 398
OY 335 SKLISNGSKINCIPNNCDRMFSREYDLRRH 365

```





```

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8627
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8627

Query Match      15.4%; Score 350; DB 4; Length 685;
Best Local Similarity 28.8%; Pred. No. 2,4e-24;
Matches 110; Conservative 60; Mismatches 160; Indels 52; Gaps 19;

QY 6 ETKSISLSSSSSPKKYICTYEGODKAYNRPILLQHLRTHSNDRPYKCTVDDCKA 65
DB 307 QNSLKKQKSHSEKRP--YEC--NECGKAFRRSSNLIOHRIHSEKPYVCS--ECGKA 360
QY 66 FFRKSHLETHIVHSEKRPFCGCGVNSROHLKRNHEITHT--KSFKCTFENCODAFY 123
DB 361 FRRSSNLIKHNRHTHGEKPECEGCGKAFSGSAHLKHQVHTHGEKPYEC--NDCKGPRS 418
QY 124 KHQSL-RHHLIVSEKTLTKCKCNKVTFRPSKLAQHLKHHGSPAYQCDHPGCFKPNFOT 182
DB 419 RVSNLILKHHRVHTGKPKYKCDGKAFSGSSSLIOHRIHTGKRP--HVCN--VCGKAFSY 475
QY 183 WSLYQPH--IKQSHPKLCKPKCGKCGVKGKGLSHMLSHDSTMIKTIWTCDYDVGR-FA 239
DB 476 SSVLRKQIHTHGEKPRCSVCGKAFSSSALLIOHGVH--TGDPYACHEC--KTFG 530
QY 240 KKNELVEHYNIIFHDGNIPIDDLKETEYKLENLIDQSKLNLHLETEKLYEDEDDE 299
DB 531 RSNLILHQRV-HTGEKPYCTGCKTFGSSSTLIOHRIH-----GLKHECNGC- 581
QY 300 EDSLDEKRSQVRSDSMSAQRSI---KSFTASLEGSKVS-----KLISNGSKKINC 347
DB 582 -----GKAFRRSSNLIOHRIHHTGKPYTC-VECGKFGSSSLIOHRIHTHGEKPYKC 634
QY 348 PKNCCDMFSEYDLRHLKMH 369
DB 635 --SECGKAFSQRSLIOHRIH 654

RESULT 11
US-09-949-016-9434
; Sequence 9434, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9434
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9434

Query Match      15.4%; Score 344; DB 4; Length 639;
Best Local Similarity 27.5%; Pred. No. 7.9e-24;
Matches 98; Conservative 47; Mismatches 129; Indels 82; Gaps 15;

QY 23 KKYICTYEGODKAYNRPSSLLEQHLRTHSNDRPYKCTVDDCKAFRRKSHLETHIVHSEK 82
```

```

DB 356 KPEYCIQ--CGKAHGHKALTDHLRIHTEKEYECA--ECGKTFRRSSNLIOHRSHTGE 411
QY 83 KPFHSGVCGKGVNSROHLKRNHEITHTKSPCTFENCQEAFFYHQSRLHLSVH--EKL 140
DB 412 KPEYCECKGSPFRYNSLLEHVRTHGEIPEYCEKCGKAFKYSSTLTGK-MRIHTEKPR 470
QY 141 TCKQCNKVTFRPSKLAQHLKHHGSPAYQCDHPGCFKQFOTWSVLOPHIK--QSHPKLK 198
DB 471 ECGECKAFSPKSKSHLIIHQRTHTKEP-YKCNB--CGKAFGSSSLITMRRHTHGSPP 527
QY 199 CPKCGKGVCGKGLSMLSHDSTMIKTIWTCDYDVGR-PAKKNELVEHYNIIFHDGNI 257
DB 528 CNOGCGKFGRIQBLTGQHRVH--TGKPYECNEC--GKAFSGSHLIVHQRHT-FTG--- 578
QY 258 DDLKETEYKLENLIDQSKLNLHLETEKLYEDEDDEDSLDEKRSQVRSDSMSA 317
DB 579 -----EKPYECNECEKAFNAKQVLVH 600
QY 318 ORSKSFTASLEGSKVSKLISNGKK-INCPNNCDMFRSEYDLRHLKMHDN 372
DB 601 QR-----SHTEKPYEC--NECGKTFKQNASLITKIVKTHSED 635

RESULT 12
US-09-949-016-11494
; Sequence 11494, Application US/09949016
; Patent No. 6812338
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11494
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11494

Query Match      15.4%; Score 343; DB 4; Length 746;
Best Local Similarity 27.3%; Pred. No. 1.2e-23;
Matches 118; Conservative 65; Mismatches 141; Indels 108; Gaps 22;

QY 23 KKYICTYEGODKAYNRPSSLLEQHLRTHSNDRPYKCTVDDCKAFRRKSHLETHIVHSEK 82
DB 336 KPFKCT--ECGKAFRRSSNLIOHRIHTEKPYRC--EECGKAFGSSSLIHHQRHTGE 391
QY 83 KPFHSGVCGKGVNSROHLKRNHEITHT--KSFCTFENCODAFYKHQSLNH----- 131
DB 392 RPYGCREGCKAFSGQSLVHQRHTHGERPYPC--KECGKAFSGSSSLIOHRIHHTGEKA 449
QY 132 -ILS-----VH--EKLTKQCNKVTFRPSKLAQHLKHHGSPAYQCDHPG 176
DB 450 QILKASDSPSLVAHQHRIHVERKPKDCGKARWISRLSLOHLITGKRP-YKCNK--C 506
QY 177 FKNFOTWSVLOPHIKOSH--PKLCKPKCGKGVCGKGLSMLSHD--STMIKTIWTCDY 232
DB 507 TKAFGSSSLIIR-QTHHTEKPEKFCDEGCKGFV-----QSHLIIHQRIHTEKRPVND 561
QY 233 CDVGR-PAKKNELVEHYNIIFHDGNIPIDDLKETEYKLENL-----DOGSFLANL 282
DB 562 C--GKAFSGSSSLIYHQRH-HKEKPYECILOGCKAFSMTOLTIIHQHVHTHGERPYCNBC 618
```

```
QY 283 -----HELETEKLVDEDE-----DEEDSLDEKRSDVRSMSAQSISKP 324
Db 619 GKAFQNSSTLFOHQIHAGVPECECGKAFSSSYLIEHQIHTR-----QWPEYV 672
QY 325 TASLEGSYSKXLSNSGKKI-----NCPK-----NNCPRMS 357
Db 673 GNALESTVSRKKVNTIKKHOCEDECKI FMRSHLIIHQIHTEKPYKCDGKAFN 732
QY 358 REYDLRRHLKWH 369
Db 733 RSSRLTQHOKIH 744

RESULT 13
US-09-949-016-8149
; Sequence 8149, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8149
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8149

Query Match 15.4%; Score 343; DB 4; Length 769;
Best Local Similarity 32.2%; Pred. No. 1.3e-23;
Matches 86; Conservative 35; Mismatches 100; Indels 46; Gaps 11;

QY 23 KKIYCTYEGCDKAYNPSSLLEOHLRTSHNDPRYKCTVDDCDKAFPRKSHLETHIVSHSR 82
Db 516 KPYKC--EECGKQFSSASSFSQSHQVHTGEKPRCNV--CGKQFSSSYQAHQVHTGE 571
QY 83 KPHGCGVCGKGVNSRQHLKRNHEITHT--KSFCTFENCQDAFYKHOSLHHILSVH--EK 138
Db 572 KPYKCEVCGRRFWSLNLNHHQRVHTGEKPYKC--EECGKQFSSQASNLQAH--QSVHTGEK 628
QY 139 TLTCQCNKVFTPRPSKLAQHLKHHGSPAYOCDDHPCFPNPTQWVLOPH--IKQSHPK 196
Db 629 PRKCDACQKRFSSQASHLQAHQVHTGEKPR--YKCD--TCGKAFSSQNSNLQVHQIHTEKPR 685
QY 197 LKCPKCGKCGVCGKGVNSRQHLKRNHEITHT--KSFCTFENCQDAFYKHOSLHHILSVH--EK 236
Db 686 FKCEBEGKESWAGSAGSAHQRVH---TGKPYTCQCGKQFSSQASHFHTHQVHTGERPY 742
QY 237 -----KFAKKNELVEHNIPIFDGNI 256
Db 743 ICDVCCGKQFSSQASHLQAHQVHTGNTL 769

RESULT 14
US-09-538-092-1275
; Sequence 1275, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
```

```
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurataseqFormatter Version 0.9
; SEQ ID NO 1275
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q06730
US-09-538-092-1275

Query Match 15.4%; Score 343; DB 4; Length 810;
Best Local Similarity 28.3%; Pred. No. 1.4e-23;
Matches 112; Conservative 57; Mismatches 153; Indels 74; Gaps 20;

QY 18 SSSRPKTYICTYEGCDKAYNPSSLLEOHLRTSHNDPRYKCTVDDCDKAFPRKSHLETHIV 77
Db 377 SHTGEKPEEC--NECGKAFSHKSALTLMQHTHTEKPYQCNH--CGKTFQCKSDLTQQR 432
QY 78 SHSEKKPFCVCGKGVNSRQHLKRNHEITHT--KSFCTFENCQDAFYKHOSL--RHILLS 134
Db 433 THTGAKPYCEYECGSPFVTSILKXHQHTHTEKPECL--ECGKSFSEKSLTQHQRH 490
QY 135 VHEKTLTKQCNKVFTPRPSKLAQHLKHHGSPAYOCDDHPCFPNPTQWVLOPHIKOSH 194
Db 491 IGDKSEYCNACGKTLYHNSLTLRHQIHTEKWP--YECSE--CGKTFCKSDLTVA--QRTH 546
QY 195 PVLK---CPKQSGKGVCGKGVNSRQHLKRNHEITHT--KSFCTFENCQDAFYKHOSL--RHILLS 250
Db 547 TGQKPFACBEGKQFSSQASHLQAHQVHTGEKPRCNV--TGKPYTCQCGKQFSSQASHFHTHQVHTGERPY 742
QY 251 FHDGNIIPDDLKETEVEKLENLDOGSXYLNLNHELETEKLVKDEDE-----D 298
Db 601 THTGE-----KPYCNCQKAFYQKSL-----TQHRIHTEKPYKCNBEGKARQCH 647
QY 299 BEDSLDEKRSDVRSMSAQSISFTASL-----EGSKV-----SKL 337
Db 648 KSALIVHQRTHQEKPYKCNBEGKSFVCKSGLIPIERKHTGKPYECNCGKFFPKKSL 707
QY 338 I-----SNSGKTKNCNNCDRMSREYDLRRHLKWH 369
Db 708 TVHHRATHTEK--SCQCNCGKIFPKKSLAQHQRSH 742

RESULT 15
US-09-538-092-1195
; Sequence 1195, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurataseqFormatter Version 0.9
; SEQ ID NO 1195
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
```

; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P52742  
US-09-538-092-1195

Query Match 15.1%; Score 337; DB 4; Length 469;  
Best Local Similarity 28.2%; Pred. No. 2.3e-23;  
Matches 103; Conservative 49; Mismatches 163; Indels 50; Gaps 17;

QY	23	KKYICTYEGGDKAYNRP	SLLEQHLRT	HSNDRPYKCTV	DDCKAFPRKSHLETIV	SHSEK	82
DB	79	KPYKCTQ--CGRTFNQI	APLIQHQR	HTHGKPYECS--	ECGKSFSPRSSP	QHERHTHG	134
QY	83	KPFHSCVCGKGVNSR	OHLEKHEITHT--	KSFCTFENCCEAFYKH	OSL--RHHILSV	HEKT	139
DB	135	KPYEGCEGCGKAFR	QSIHLTQHRL	IHTGEKPYQC--	GBCGKAFSHSS	SLTQHRIHTGEKPY	192
QY	140	LTCGQCNKVFT	RPSKLAQHKLKH	HGSPAYQCDHPGCF	KNFTWVLIQFH--	IKQSHPKL	197
DB	193	YECHEGCKAFTQIT	PLIQHQR	HTHGKPY--YECGE--	CGKAFSOSTLL	TEHRR	249
QY	198	KCPKCGKGCYGGK	GLSHMLSHD	STMIKIWTG	CDVGK--FAKKNEL	VEHVNIPHDGNI	256
DB	250	GCNECGKTF	SHSSLSQHER	TH--TGEKPYECSQC--	GKAFRQSTH	LTQHQR	302
QY	257	PDDLKETEYVK	LENLDDGSK	LNINLHELETEK	LKYEBDE-----	EDEEDSLD	304
DB	303	-----KPYECND	CGKAFSHSS	SLTQHRIHTHGKPYE	CNCGRAFSQ	LAPLIQHQR	357
QY	305	EKRSDVRS	DSMSAQR	SIKSF	TASLEGS	KSVSKLISNGKK	364
DB	358	EKPYECNQC	GRASARA---	TLLIEHQR-----	IHTGEKPYQC--	NBCGKSF	406
QY	365	HLKMH	369				
DB	407	HERTH	411				

Search completed: February 8, 2005, 15:23:40  
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:22:12 ; Search time 129 Seconds  
(without alignments)  
1040.228 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229

Sequence: 1 MESSDETKSISLSISSSSSS.....PLYKARMDLPNFTVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1373511 seqs, 325702437 residues

Total number of hits satisfying chosen parameters: 1373511

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/prodata/2/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/2/pubppa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/2/pubppa/US06\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/2/pubppa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/prodata/2/pubppa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/2/pubppa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/prodata/2/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/prodata/2/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/prodata/2/pubppa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/prodata/2/pubppa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/prodata/2/pubppa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/prodata/2/pubppa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/prodata/2/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/prodata/2/pubppa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/prodata/2/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/prodata/2/pubppa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/prodata/2/pubppa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/prodata/2/pubppa/US10\_NEW\_PUB.pep:\*  
19: /cgn2\_6/prodata/2/pubppa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/prodata/2/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2225	99.8	412	US-10-032-585-7493	Sequence 7493, Ap
2	445	20.0	564	US-10-128-714-3313	Sequence 3313, Ap
3	445	20.0	564	US-10-128-714-3313	Sequence 3313, Ap
4	400	17.9	404	US-10-108-260A-4601	Sequence 4601, Ap
5	398	17.9	487	US-10-029-386-33849	Sequence 33849, A
6	397.5	17.8	423	US-10-734-564-83	Sequence 83, Appl
7	395.5	17.7	409	US-10-106-698-5955	Sequence 5955, Ap
8	395.5	17.7	1191	US-10-408-765A-2105	Sequence 2105, Ap
9	392.5	17.6	595	US-10-203-052B-8	Sequence 8, Appl
10	392	17.6	555	US-10-094-749-2331	Sequence 2331, Ap
11	391	17.5	552	US-10-108-260A-3669	Sequence 3669, Ap
12	390	17.5	540	US-10-104-047-3748	Sequence 3748, Ap
13	390	17.5	576	US-10-094-749-2365	Sequence 2365, Ap

14	388	17.4	568	US-10-203-052B-7	Sequence 7, Appl
15	387.5	17.4	742	US-10-029-386-31985	Sequence 31985, A
16	387.5	17.4	839	US-10-408-765A-1850	Sequence 1850, Ap
17	385	17.3	577	US-10-381-327-3	Sequence 3, Appl
18	378	17.0	613	US-10-408-765A-1517	Sequence 1517, Ap
19	376	16.9	499	US-10-108-260A-2954	Sequence 2954, Ap
20	374	16.8	566	US-10-322-774-10	Sequence 10, Appl
21	373	16.7	452	US-10-471-450-5	Sequence 5, Appl
22	372.5	16.7	586	US-10-104-047-2592	Sequence 2592, Ap
23	372	16.7	393	US-09-911-261A-12	Sequence 12, Appl
24	372	16.7	393	US-10-057-408-12	Sequence 12, Appl
25	372	16.7	393	US-10-333-487-12	Sequence 12, Appl
26	371	16.6	403	US-10-029-386-32933	Sequence 32933, A
27	370	16.6	530	US-10-094-749-2723	Sequence 2723, Ap
28	369.5	16.6	464	US-10-108-260A-3590	Sequence 3590, Ap
29	367	16.5	514	US-10-029-386-34112	Sequence 34112, A
30	367	16.5	610	US-09-989-920-212	Sequence 212, App
31	367	16.5	641	US-10-363-616-421	Sequence 421, App
32	364.5	16.4	444	US-10-108-260A-3223	Sequence 3223, Ap
33	364	16.3	516	US-09-974-298-118	Sequence 118, App
34	364	16.3	535	US-10-264-048-3016	Sequence 3016, Ap
35	363	16.3	516	US-10-094-749-1976	Sequence 1976, Ap
36	361	16.2	803	US-10-408-765A-1049	Sequence 1049, Ap
37	360	16.2	529	US-10-108-260A-3085	Sequence 3085, Ap
38	360	16.2	791	US-10-471-450-18	Sequence 18, Appl
39	360	16.2	924	US-10-007-280A-231	Sequence 231, App
40	359.5	16.1	365	US-10-094-749-2297	Sequence 2297, Ap
41	359.5	16.1	427	US-10-755-889-370	Sequence 370, App
42	358.5	16.1	473	US-10-029-386-32047	Sequence 32047, A
43	358	16.1	406	US-09-764-864-976	Sequence 976, App
44	358	16.1	457	US-10-377-079-68	Sequence 68, Appl
45	358	16.1	647	US-10-377-079-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1				
US-10-032-585-7493				
Sequence 7493, Application US/10032585				
Publication No. US20030180953A1				
GENERAL INFORMATION:				
APPLICANT: Terry, Roemer D.				
APPLICANT: Bo, Jjiang Boone				
APPLICANT: Howard, Bussey				
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery				
FILE REFERENCE: 10182-005-999				
CURRENT APPLICATION NUMBER: US/10/032, 585				
CURRENT FILING DATE: 2001-12-20				
NUMBER OF SEQ ID NOS: 8000				
SOFTWARE: PatentIn version 3.1				
SEQ ID NO 7493				
LENGTH: 412				
TYPE: PRT				
ORGANISM: Candida albicans				
US-10-032-585-7493				
Query Match				
Best local Similarity 99.8%; Pred. No. 2, 3e-154;				
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	MSSDETKSISLSISSSSSRPKKYCTYEGCDKAYNPBLLFOHLRTHSNDRPYKTV	60	
Db	1	MSSDETKSISLSISSSSSRPKKYCTYEGCDKAYNPBLLFOHLRTHSNDRPYKTV	60	
Qy	61	DCCKAFRRSHLETHVSHSEKRPFCVCGKVNRRHDKHREITHTSFKTFENCOR	120	
Db	61	DCCKAFRRSHLETHVSHSEKRPFCVCGKVNRRHDKHREITHTSFKTFENCOR	120	
Qy	121	AFYKHSLSRHHLSVVEKTLTKCKCNKVFTRPSKLAQHLKHGSPAYGCDHPGCFKNF	180	
Db	121	AFYKHSLSRHHLSVVEKTLTKCKCNKVFTRPSKLAQHLKHGSPAYGCDHPGCFKNF	180	



```
Qy 181 QWTVLQFHIIKOSHPRKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVAKFAK 240
| | | | |
Db 181 QWTVLQFHIIKOSHPRKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVAKFAK 240
Qy 241 KNEIVHENYI FHDGNI PDDLKETEYKLENI LDOSKXNNIHELETEKXKVEDEDEDE 300
| | | | |
Db 241 KNEIVHENYI FHDGNI PDDLKETEYKLENI LDOSKXNNIHELETEKXKVEDEDEDE 300
Qy 301 DSIDEKRSVDVSDMSAORSIKSFTASLEGSKSXSLISNSGKKNICPKNCDRMFSREY 360
| | | | |
Db 301 DSIDEKRSVDVSDMSAORSIKSFTASLEGSKSXSLISNSGKKNICPKNCDRMFSREY 360
Qy 361 DLRRLHKWHDNIQRIESFLNLSIEKEETPEGEPLVKKAMDLLPNETSVISR 412
| | | | |
Db 361 DLRRLHKWHDNIQRIESFLNLSIEKEETPEGEPLVKKAMDLLPNETSVISR 412
```

## RESULT 2

```
US-10-128-714-3313
/ Sequence 3313, Application US/10128714
/ Publication No. US20030119013A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Bo
/ APPLICANT: Hu, Wengqi
/ APPLICANT: Tishkoff, Daniel
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Broshkin, Alexey M
/ APPLICANT: Lemieux, Sebastien M
/ TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
/ FILE REFERENCE: 10182-018-999
/ CURRENT APPLICATION NUMBER: US/10/128,714
/ PRIOR FILING DATE: 2002-04-23
/ PRIOR APPLICATION NUMBER: US 60/285,697
/ PRIOR FILING DATE: 2001-04-23
/ PRIOR APPLICATION NUMBER: US 60/287,066
/ PRIOR FILING DATE: 2001-04-27
/ PRIOR APPLICATION NUMBER: US 60/295,890
/ PRIOR FILING DATE: 2001-06-05
/ PRIOR APPLICATION NUMBER: US 60/303,899
/ PRIOR FILING DATE: 2001-07-09
/ PRIOR APPLICATION NUMBER: US 60/316,362
/ PRIOR FILING DATE: 2001-08-31
/ NUMBER OF SEQ ID NOS: 8603
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3313
/ LENGTH: 564
/ TYPE: PRF
/ ORGANISM: Aspergillus fumigatus
US-10-128-714-3313
```

Query Match 20.0%; Score 445; DB 14; Length 564;

Best Local Similarity 30.3%; Pred. No. 3.9e-24;

Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

```
Qy 14 ISSSSSRP---KKYICTYEGCDKAYNRPSLLEOHLRTSHNDRPYKCTVDDCDKAFPRKS 70
| | | | |
Db 71 ISTSAKYPSELKTHNCPDPCGTAKFNRPARLOEHLRSHNNERIKCTFEEBCDKTFLRAS 130
Qy 71 HLETHIVS-HSEKKPFHCS--VCGKGVNSROHLKHEITH--TKSFECT-FENCOBAFYK 124
| | | | |
Db 131 HNNHHIKSAHTGVDRDYCDRPGCGSFVTGSRRLRHLAHDGRDXYRCTEYPPCNETFRK 190
Qy 125 HOSLRHNLISVH--EKTILTC-----KQCKNVFTRPSKLAOHLKHHGSPAYOCHN- 173
| | | | |
Db 191 HSTLOKHLMTAHLKQKRPQCSPHTDPSGTQCTMAFDTAGHLRAHESRIH-TEKRSCTEC 249
Qy 174 ----PGCFKNQOTWVSLQFHIKOSHPRKLCPRKCGKGVGKGLSSHM-LSHDDSTM--IK 226
| | | | |
Db 250 SQHAGGAATFTPYALLQAHIRSVNRP-QCPNICALTCATSRLRLHLEVAHAGDVSLBERK 308
Qy 227 IWTCDY--CDVGFKAQKNEIVHENYI FHDGNI-----PDDLKETEYKLENI LDQ---- 275
```

```
Db 309 IFFCTVPGCD-RSFTKGNLTVHIRTVHOGEKRFVCGETDLSSSKKVSGMNN--DNGCGK 365
| | | | |
Qy 276 --GSKLNNLHELETEKXKVEDEDEDEBDSL-----DEKRSVDVSDMSAORSIKSFTASLE 329
| | | | |
Db 366 RYGSKLALAEHIRTVAHLGYQNAKARRORLGIITRRORIGSTATSPGVA-----LAALT 418
Qy 330 GSKSVSKLISNSGKKNICPKNCDRMFSREYDLRHL--KWDNIQRIESFL 380
| | | | |
Db 419 GEGVAE-----TGRHIACTVESCPRHFRHDYDLVWMSGKHFSSEBTRDLFL 467
```

## RESULT 3

```
US-10-128-714-8313
/ Sequence 8313, Application US/10128714
/ Publication No. US20030119013A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Bo
/ APPLICANT: Hu, Wengqi
/ APPLICANT: Tishkoff, Daniel
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Broshkin, Alexey M
/ APPLICANT: Lemieux, Sebastien M
/ TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
/ FILE REFERENCE: 10182-018-999
/ CURRENT APPLICATION NUMBER: US/10/128,714
/ PRIOR FILING DATE: 2002-04-23
/ PRIOR APPLICATION NUMBER: US 60/285,697
/ PRIOR FILING DATE: 2001-04-23
/ PRIOR APPLICATION NUMBER: US 60/287,066
/ PRIOR FILING DATE: 2001-04-27
/ PRIOR APPLICATION NUMBER: US 60/295,890
/ PRIOR FILING DATE: 2001-06-05
/ PRIOR APPLICATION NUMBER: US 60/303,899
/ PRIOR FILING DATE: 2001-07-09
/ PRIOR APPLICATION NUMBER: US 60/316,362
/ PRIOR FILING DATE: 2001-08-31
/ NUMBER OF SEQ ID NOS: 8603
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 8313
/ LENGTH: 564
/ TYPE: PRF
/ ORGANISM: Aspergillus fumigatus
US-10-128-714-8313
```

Query Match 20.0%; Score 445; DB 14; Length 564;

Best Local Similarity 30.3%; Pred. No. 3.9e-24;

Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

```
Qy 14 ISSSSSRP---KKYICTYEGCDKAYNRPSLLEOHLRTSHNDRPYKCTVDDCDKAFPRKS 70
| | | | |
Db 71 ISTSAKYPSELKTHNCPDPCGTAKFNRPARLOEHLRSHNNERIKCTFEEBCDKTFLRAS 130
Qy 71 HLETHIVS-HSEKKPFHCS--VCGKGVNSROHLKHEITH--TKSFECT-FENCOBAFYK 124
| | | | |
Db 131 HNNHHIKSAHTGVDRDYCDRPGCGSFVTGSRRLRHLAHDGRDXYRCTEYPPCNETFRK 190
Qy 125 HOSLRHNLISVH--EKTILTC-----KQCKNVFTRPSKLAOHLKHHGSPAYOCHN- 173
| | | | |
Db 191 HSTLOKHLMTAHLKQKRPQCSPHTDPSGTQCTMAFDTAGHLRAHESRIH-TEKRSCTEC 249
Qy 174 ----PGCFKNQOTWVSLQFHIKOSHPRKLCPRKCGKGVGKGLSSHM-LSHDDSTM--IK 226
| | | | |
Db 250 SQHAGGAATFTPYALLQAHIRSVNRP-QCPNICALTCATSRLRLHLEVAHAGDVSLBERK 308
Qy 227 IWTCDY--CDVGFKAQKNEIVHENYI FHDGNI-----PDDLKETEYKLENI LDQ---- 275
| | | | |
Db 309 IFFCTVPGCD-RSFTKGNLTVHIRTVHOGEKRFVCGETDLSSSKKVSGMNN--DNGCGK 365
| | | | |
Qy 276 --GSKLNNLHELETEKXKVEDEDEDEBDSL-----DEKRSVDVSDMSAORSIKSFTASLE 329
| | | | |
Db 366 RYGSKLALAEHIRTVAHLGYQNAKARRORLGIITRRORIGSTATSPGVA-----LAALT 418
```

QY 330 GSKSVSKLISNGKKINCPKNCNDKMFSSREYDIDRLH--KWHDNLTQRIESEL 380  
| : | : | : | : | : | : | : |  
Db 419 GEGVAEE---TGRHIACLVESCPHRFHRDYDLVWMSGKHFFSEEEERDLFL 467

```

RESULT 4
US-10-108-260A-4601
; Sequence 4601, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4601
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4601

```

Query Match	17.9%	Score 400;	DB 15;	Length 404;
Best Local Similarity	33.9%	Pred No. 5e-21;		
Matches 121;	Conservative 39;	Mismatches 163;	Indels 34;	Gaps 19;

Qy	23	KYUITYEBCDZYUNRPSLLEOHLRTHSNDRPKVCVUDODKAFPKRSHLTHETIVSHSEK	82
Dp	39	KPYKC--BECGKAFNRSSILTKHKIITHGEKPYC--BECGKGFSSVTLNTHKAIHABE	94
Qy	83	KPHCSVCGKGVMSKROHLKRNHETHT--KSFCTFENCQEAFLKQSLRNHILSVH--EK	138
Dp	95	KPYKCEBGKASNNSSSKIMEHKRIHNGEKPYPKC--BECGKAFSSSLTEH--RRIHAGEK	151
Qy	139	TLTKQCNKVPFRPSPLAOKHLKHHGSSPAYOCSDHPCEPNNFQVTLWLPFH--IKOSHPK	196
Dp	152	PKYCEBGCAFNRSSILTKHKIITHGEKP--KYCE--GGCAKAFSVSTLNTNKAIIHABEK	208
Qy	197	LKPRCKGKGVCKGKGLSSHMLSHDSBTMIKIWTCDYCDVQK--PAKQELVEHNIITDGN	255
Dp	209	KYCEBGKASNNSSSKIMEHKRIH--TGKRPKYCEBC--GKASWSSSLTEHRI--HAGE	262
Qy	256	IPDDILKETEVKLLLENILDOGSSKANNLHLETEKLVYEE--DEEDEDSLDEKSDVRSDS	314
Dp	263	KP---KYCEBGKAFWSSSFLTKRKIIHAAE--KPYCEBGKGFSPFSLITKHKIITHGE	318
Qy	315	--MSQRSTKSTTASILEGSKSVKSLISNGSKKINCPYRANNCNDRMFSEYDLRRLKYN	369
Dp	319	KRYKCEBGKAF--SWSSILTEHKIITHGEKPYC--BECGKAFSSSSSLTKRKRIH	371

```

RESULT 5
US-10-029-386-33849
: Sequence 33849, Application US/10029386
: Publication No. US20030194704A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
: TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
: FILE REFERENCE: AEOICA-X-2
: CURRENT APPLICATION NUMBER: US/10/029,386
: CURRENT FILING DATE: 2001-12-20
: NUMBER OF SEQ ID NOS: 34288
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 33849
: LENGTH: 487
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:

```

```

; OTHER INFORMATION: MAP TO AC011467.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
; OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUUE 0.00e+00
US-10-029-386-33849

```

Query Match	17.9%	Score 398	DB 14	Length 487
Best Local Similarity	30.2%	Pred. No. 8.7e-21		
Matches 121; Conservative	48;	Mismatches 167;	Indels 64;	Gaps 18;

QY	20	RRPKYIITYGCDXANRPSLLEOHLTPHSNDRPKYCDVDDODKAFPRKSHLETHVSH	79
Db	120	SGEKPYC--YECGKANVETSNLSHTKRIHTGKKPKYC--BEGGKAFNLSHTLTKIITH	175
QY	80	SEKRPHCSCGVNSRQHLRHEIHTY--KSFKTEPENCCEAFPKQSL--RHHLISVH	136
Db	176	TGKKRPYKCEEGKANQSANLTHKRIHTGEKPYC--BEGGKAFQSSLTPLHAKI IHG	233
QY	137	EKTLTCKQCNVFTFRPSKLAQHLKHNHGSPAYQCDHPCCFPNFQTSVLQDH--IKQSH	194
Db	234	EKPVCCECGKAFQSSLTTHKI--IHGGEKPYCEE--CGKAFQSLSHLTHKRIHSG	290
QY	195	PKLKCPKRGKGVCKGKGLSSHNLSHDDSTMILMTCVQDYDVKFKAKNLEVHYNIFHG	254
QY	291	KPYKCEEGKAFKQSSLTTHKRIHAG--KPYCEVCSKA--FSKPSHLTHKRI--HTG	345
QY	255	NIPDDLKETEYKLENLDDQSKLNNLHELETKYKVEDEDEDEDSLDEKRDVRSBS	314
Db	346	EKP--YCEBEGKAFNLSQULTHKI IHTGE--KPYKCEC-----GKAFNQST	391
QY	315	MSAQRSI-----KSFTASLEGSKSVSKLINSKGKINCPCNNCDMPFSREYD	361
Db	392	LSKHV IHTGEKPYKYCEGKAFNQS--SHLTTHKMIHTGEKPYC--EECGKAFNNSI	447
QY	362	LRRLHMHND-----DLQRIESTLNSIEKE	387
Db	448	LNRRHMIHTGEKLYKPSCNACNADIAKISKYKRCAGGK	487

```

RESULT 6
US-10-734-564-83
; Sequence 83, Application US/10734564
; Publication NO. US2004015728A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FASTSEQ for Windows Version 4.0.
; SEQ ID NO 83
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-83

```

	Query Match	Similarity	17.8%	Score 397.5	DB 16	Length 423
	Best Local Match	31.1%	Pred. No. 8e-21			
	Matches	99	Conservative	55	Indels	69
						Gaps 16
Qy	8	KSISL-----ISSSSSR-----PKYICTYEGCDKAYRPSLLEOHLRTHSNDR	53			
Db	67	ESVESLITADAFIAAGESSAPTPRPALPRPFIISFPDCSANYSAWMLDLHLCHTGER	126			
Qy	54	PKCTVDDCDKAFPRKSHLETHIYVSHSEKPEFHSV--CGKGVNSRQHLKRN-ELTH--	107			
Db	127	PVQVDYBGCAGAFRDYHLSRHILTHGEXKPEVANAAGCCDQKFTYKSLKGFERKHNQ	186			
Qy	108	TKSEKCTFENCQAEAFYKHOSLR-HLILSVHEKTLTCKQ--CNKFTPRSKLAQHLKHG	164			
Db	187	QKQYICSEDECDKTFPKKHQDLAKHQCNQTNRPFLPKCTQEGCGKFAFSISKLRRAKHAHG	246			
Qy	165	GSPAYQCDHPCGCFKNNFQTVSLQFHLKOSHDK-----	196			

```
Db      247 ---YVC---OKGCSFVAKTTELLKRETHKEIILCEVCRKTFKRDYIKQHNKTHAPER 301
      197 --LKCPK--CGKGCYVKKGKGLSSHMLS-HDDSTMIKWTCDYCVGK-PAKKNELVEHYNI 250
      302 DVCRCRCEGCGRTYTTVFNLQSHILSFHES---RPFVEHAGCGCTFAMKQSLTTH-AV 357
Qy      251 FHDGNI PDDLKETEYK 268
      358 VHD---PDKKKMKLKYK 372

RESULT 7
US-10-106-698-5955
; Sequence 5955, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5955
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5955

Query Match      17.7%; Score 395.5; DB 14; Length 409;
Best Local Similarity 31.1%; Pred. No. 1.1e-20;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

Qy      8 KAISLSL-----ISSSSSR-----PKYICTYEGCDKAYNRPSSLLEOHLRTSHNR 53
      53 ESVSLITADAPFLAAGESSAPTPPRALPRRPFICSPPDGSANYSKAMKDAHLCKHTGER 112
Qy      54 PKCTVDDDDKAFPRKSHLETHIVSHSEKKPRHCV--CGKGVNSRQHLKRH-ETH--- 107
      113 PFCVCEGCGKAFIRDYHLSRHILTHTGKPRVCAMGCDOKPNTKSNLKHFERRGHENQ 172
Qy      108 TSPFKCTPENCCEAFYKHOSLR-NHILSVHEKTLTCKO--CNKVFRPSKLAQHLKXHG 164
      173 QKQYISFSDCKKTKFKHQQLKHOCQHTNEPLFKCTQCGCGHGFSPSKLKHAKAHG 232
Qy      165 GSPAYOCDHPGCFKFOFWSVLQFHIKOSHPRK----- 196
      233 ---YVC---OKGCSFVAKTTELLKRETHKEIILCEVCRKTFKRDYIKQHNKTHAPER 287
Qy      197 --LKCPK--CGKGCYVKKGKGLSSHMLS-HDDSTMIKWTCDYCVGK-PAKKNELVEHYNI 250
      288 DVCRCRCEGCGRTYTTVFNLQSHILSFHES---RPFVEHAGCGCTFAMKQSLTTH-AV 343
Qy      251 FHDGNI PDDLKETEYK 268
      344 VHD---PDKKKMKLKYK 358

Db      344 VHD---PDKKKMKLKYK 358

RESULT 8
US-10-408-765A-2105
; Sequence 2105, Application US/10408765A
; Publication No. US20040101874A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Colin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2105
; LENGTH: 1191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2105

Query Match      17.7%; Score 395.5; DB 16; Length 1191;
Best Local Similarity 35.8%; Pred. No. 4e-20;
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

Qy      20 SRPKYICTYEGCDKAYNRPSSLLEOHLRTSHNDREPKYCTVDDCDKAFPRKSHLETHIVSH 79
      905 TREKTYKC--EECGKAFQPSHLLTHHKMHTGEKEYKC--EECGKAFQSSSTLTHKILH 960
Qy      80 SEKRPFGSVCGKGVNSRQHLKRHEITHT--KSPFCTENCOEAYVQISL-RHILLSVH 136
      961 TGEKPKCEBCKGKAFRKSSTLTENKIHTGEKPYKC--EECGKAFQSSSTLTHRHRTMG 1018
Qy      137 EKLTCCKCNKVFTPSPKLAQHLKHHGSPAYOCDFNPFOTWGLQFHIKOSHPRK 196
      1019 EKPYKCEBCKGAFNRSSLTTHKIHITGEKP-YKCE--CGKAFSSSTLNGH-KRIHTR 1074
Qy      197 --LKCPKCGKGVCKKGLSSHMLSDDSTMIKWTCDYCVGKPAKKNELVEHYNIFPD 253
      1075 EKPYKCEBCKGAFQSSSTLTHKRLH---TGEKPYKCEC--GKAFKSSALTCKKIHT 1129
Db      1075 EKPYKCEBCKGAFQSSSTLTHKRLH---TGEKPYKCEC--GKAFKSSALTCKKIHT 1129
Qy      254 GNIPDDLKETEYKLEMLDQGSCLNLIHLET 287
      1130 GE-----KPYKCEKGAFFNQSSTLTHKRIHT 1157
Db      1130 GE-----KPYKCEKGAFFNQSSTLTHKRIHT 1157

RESULT 9
US-10-203-052B-8
; Sequence 8, Application US/10203052B
; Publication No. US20040109854A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Zang Hee
; APPLICANT: Komed CO., LTD.
; APPLICANT: Kim, Hong Hee
; TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein
; FILE REFERENCE: 59520-00002
; CURRENT APPLICATION NUMBER: US/10/203,052B
; CURRENT FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-052B-8

Query Match      17.6%; Score 392.5; DB 16; Length 595;
Best Local Similarity 31.7%; Pred. No. 2.8e-20;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

Qy      10 ISSLSISS--SRPKYICTYEGCDKAYNRPSSLLEOHLRTSHNDREPKYCTVDDCDKAFPR 68
      186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTHKRIHITCEKPYKC--EECGKAFNQ 241
```

[illegible]

```

RESULT 10
US-10-094-749-2331
; Sequence 2331, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: MAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2331
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2331

Query Match      17.6%; Score 392; DB 15; Length 555;
Best Local Similarity 32.5%; Pred. No. 2.8e-20;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17,

Cy      21  RPKKYICITCEGCDKAYNPNPDLLEQHLRTHSNDRPKCTVDDCCDAFFPKSHLEHTYVSHS 80
Db      169  RENSVCQ--EECDCKVFKKPSLTTRKRVHTGEKPKC--EECGKAFKHSSTLTTHKMHT 224

```

```

0Y      81  EKKPHEGCVGCGVNSROHLKHEHTHT--KSFKTFECSOAFYUGSL-RHILISVHE 137
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      225  GEKPYRCECGKAFHSHSLTHTHKVIHGEKPFKC--EECGKAFFNPSALTHHKFLHYKE 282
0Y      138  KTLTKCGCNKVPTRPSKLAQHKLKHNGSDPAVQCDHPGCFKQFQWVSLQFH--IKOSH 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      283  KPYKEEEDCKAFNRSYLTLKHXI-DHSEKSYKCEQ--CGKGFNMSSTLTKRRIRHTGEK 339
0Y      196  KLCPECKGCGVCGKSGSHMSHDSTPMIKCIWTCDCYCVGCFPAKKNELVEHNI FPDGN 255
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      340  PYKCECGKAFNVSHTLTHKVIH--IGEKYKCEEC--GAAFNHSSKLTTHIKIHTGE 394
0Y      256  IPDDLKETEYKLENLLDQSKLNNLHELET-EKX-KVEEDEDEDESDLD--EKRSDV 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      395  KP---YKCECGKAFFN--QSSNLTLKHKI IHGEKLYKCECGKAFNRSNLTTHKRIHT 448
0Y      311  RSDMSAQRISKSPFASLEGSKSVSKLTINSQKINCPCKNQCDPMPSREBYDLRRHK 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      449  GEKPYKCECGKAFNRSNLTK--HNILHTGEKSYKC--EECGKAFNQSSTLTTHKX 501

```

```

RESULT 11
US-10-108-260A-3669
; Sequence 3669, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3669
; LENGTH: 592
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-108-260A-3669

```

Query Match	17.5%	Score 391;	DB 15;	Length 592;
Best Local Similarity	31.1%	Pred. No. 3.6e-20;		
Matches 120;	Conservative 41;	Mismatches 161;	Indels 64;	Gaps 19;

```

OY 23 KXVJCTYGCJCKAXVPRSLLEOHLNLTSHSDRYKCTVDCDDCAFFKSHLETIIVSHSEK 82
Dd 199 KPYIC--EECGKAFFYSALTTHKRIHNGEKPKYC--DKCDDAFIASSTLSKHEIITHGK 254
OY 83 KPFHSCVCGKGVNSKHOMLKRHEIIT--KSFKCTEFCNCOEAFYKHQSLRNH----- 131d
Dd 255 KPYKCEEGKAFNQSSTLTGKHKIHTGKEPKYC--EECGKAFNQSSTLTGKHKIHTGKEK 312d
OY 132 -----ILSVH-----EKLTLTCQCMKVFTPRBPKLAOHKILHNGSPAYOC 171
Dd 313 YVCEEGCAFFKYSRIILTTHKRIHTEKEPKYCKMCKCAFIASSTLSHNEFIHM-GKKHYCK 371d
OY 172 DHGEGFKMFOVMSVLFQFI--IKQSHPLKCKPCKGCGVCKGGLSHMLSHDSTMILKIWT 229
Dd 372 EE--CGKAFIASSVLTTRHNRVHTGKEPKYKCEEGCAFIASSTLSKHEIITH--TGKKPKY 426d
OY 230 CDYCDVGR-FAKKMLVEHYNI FHDGNI PDDLKETEVYKLTENLDDGSKLNNLHLET- 287d
Dd 427 CEEC--GAFFNQSSTLTGKHKI-HNGEKP---YKCEEGKAFN-----QSSSLTGKHKIHTG 477
OY 288 -EKLTVDEDEDEEDS---LDEKSDVNSDSASQSTISFPAISLEGSAVSGLJNSGK 343
Dd 478 EKPVCCEEGCAFIASSTLTGKHKIHTREKPKYKCEEGCAF--HISTHTLTKIILHTGKEK 535
OY 344 KINGCKJNCDRMFSREYDLRHLTKM 369
Dd 536 PYRC--EECGKAFNHSATLSHKKH 559

```

## RESULT 12

US-10-104-047-3748  
; Sequence 3748, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3748  
; LENGTH: 540  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3748

Query Match 17.5%; Score 390; DB 15; Length 540;  
Best Local Similarity 30.0%; Pred. No. 3,8e-20;  
Matches 107; Conservative 61; Mismatches 155; Indels 34; Gaps 15;  
QY 23 KYICTEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82  
DB 198 KLYKC--QECDRTPNQSNLTETKDYAREKPYKC--EECGKAFNQSHTLTHKIIHTGE 253  
QY 83 KEFHSCVCGGVNSRQHLKRHEITHTKSPKCTFENCQEAIFYKHQSL-RHHILSVHEKTLT 141  
DB 254 KKYCECGCKAFNQSHTLTHKIIHTGEBOPIYCEEGKAFNQSHTLTHKRIHTGEKPYK 313  
QY 142 CKQCNKVFTRPSKLAQHLKHHGSPAYOCDHGCEKFNQFWSVLOFH--IKOSHFKLKC 199  
DB 314 CECEGKAFNRSSKLTETKRIHTGEOP--YKCEB--CGKAFNRSSNLTETKRIHTGEKPYK 370  
QY 200 PKCGKGVCKKLSHMLSHDSTMTKWTCDYCDVGR-FAKNDELVEHYNIFHDGNID 258  
DB 371 KECGKAFNRSSALTYTHKRIH--TGKPYKCEB--GKAFNRSSKLTETKRIHTGEK-- 421  
QY 259 DLLKETEVKLENNLDQSKLNNLHLETEKLVKVEDE-----EDEEDSLDEKSPDRSD 313  
DB 422 ---KPYKCECGKAFNQSHTLTHKIIHTGEBOPIYKCEEGKAFNQSHTLTHKRIHTGEK 478  
QY 314 SMSAQRISFTASLEGSKSVKLSNSGKK-INCPKNNCDMPFSEHYDLRRLKWH 369  
DB 479 PYKCECGKAFNRSS--SKLTETKRIHTGEKPYKCEB--CDKAFNQSANTLTHKRIH 530

RESULT 13  
US-10-094-749-2365  
; Sequence 2365, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHIKO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOMYUKI  
; APPLICANT: NAGAHARA, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084355/0160

; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2365  
; LENGTH: 576  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2365

Query Match 17.5%; Score 390; DB 15; Length 576;  
Best Local Similarity 28.6%; Pred. No. 4.1e-20;  
Matches 119; Conservative 57; Mismatches 168; Indels 72; Gaps 17;  
QY 3 ESDETKSISSLSRSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDC 62  
DB 217 ECDKAFNQSHTLTHKIIHTREKLNBEYCEGKAFNQSHTLTHKRIHTGEKPYKC--EEC 274  
QY 63 DKAFPRKSHLETHIVSHSEKPPHSCVCGGVNSRQHLKRHEITHT--KSPKCTFENCQ 120  
DB 275 GKAFNQSHTLTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIHTGEKPYKC--EECGK 332  
QY 121 AFYKHQSL-RHHILSVHEKTLTKQCNKVFTRPSKLAQHLKHHGSPAYOCDHGCEKFN 179  
DB 333 AFNKSHTLTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIHTGEKPYKC--EECGK 389  
QY 180 FQTSVLOFH--IKOSHFKLKCPCGKGVCKGKLSHMLSHDSTMTKWTCDYCDVGR 237  
DB 390 FNOFSNLTTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIH--TGKSKYKCEB--GK 444  
QY 238 -FAKNDELVEHYNIFHDGNIDDLLKETEVKLENNLDQSKLNNLHLETEKLVKVEDE 296  
DB 445 AFYRSSKLTETKRIHTGEKPYKCEB--YTCCEGKAFNQSHTLTHKRIHTGE 490  
QY 297 EDEEDSLDEKSPDRSDSMSAQRISFTASLEGSKSVKLSNSGKKINCPKNNCDMPF 356  
DB 491 KPYQ-----CECGKAFNQSHTLTHKRIHTGEKPYKCEB--CGKAF 530  
QY 357 SREYDLRRLKWHNDLORISFLNSIKETPBGEPYVKKARNDLLENETSVISR 412  
DB 531 NQSSNLTTHKRIH-----TGKLYKPRKCNSEDNETSKSKF 566

RESULT 14  
US-10-203-052B-7  
; Sequence 7, Application US/10203052B  
; Publication No. US20040109854A1  
; GENERAL INFORMATION:  
; APPLICANT: KOMED CO., LTD.  
; APPLICANT: Lee, Zang Hee  
; APPLICANT: Kim, Hong Hee  
; TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein  
; FILE REFERENCE: 59520-00002  
; CURRENT APPLICATION NUMBER: US/10/203,052B  
; CURRENT FILING DATE: 2003-11-03  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7  
; LENGTH: 568  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-203-052B-7  
Query Match 17.4%; Score 388; DB 16; Length 568;  
Best Local Similarity 28.5%; Pred. No. 5.7e-20;  
Matches 105; Conservative 62; Mismatches 167; Indels 34; Gaps 13;  
QY 15 SSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCKAFPRKS 70

Db 212 SSKLTHKRIYTCCKLYKCOECDRTNFQSNLTYYKKDYAREKPYKC--EECGKAFNQSS 269  
QY 71 HETHIVSHSEKRPFCOSVCGKGVNSRQHLKREHITHTSFKCTFENCOAFYKHQSL-R 129  
Db 270 HLTHKIHITGKPYKCEBGKASNOFSNLTTHKHTGEOPIYCECGKAFQSSSTLT 329  
QY 130 HHILSVHEKTLTCKQCNKVFTRPSTLAQHLKHHGSPAYQCDHPGCFKXFTWVSYLQFH 189  
Db 330 HKRIHTEKPYKCEBGKAFNRSKLTENKNIHTGEPR-YKCEB--CGKAFNNSULTEH 366  
QY 190 --IKOSHPRYLKCPCKCGKGVCKKGLSSHMLSHDOSTMIKIWTCDYCDVGKFAKNELVEH 247  
Db 387 RKIHTEEKPYKCEBGKAFNRSALYTHKRIH--TGEKPYKCEB--CGKAFNRSULTE 441  
QY 248 YNIFHDGNI PDDLKETEYVKLNNLDDQSKLNNHLETEKLVBEDE----EDBEDS 302  
Db 442 YKLIHTGK-----KPYKCECGKAFQSSKLTENKNIHSGELPYKCEBGKAFKSSSL 495  
QY 303 LDEKSDVSDSMASQRIKSFYASLEGSKVSGLISNGKK--INCPKNNCDMPFREVYD 361  
Db 496 THKRIHTGKPYKCEBGKAFNRS--SKLTENKNIHTGKPYECER--CDKAFNOSAN 550  
QY 362 LRRHLKWH 369  
Db 551 LTRGKKKH 558

RESULT 15  
US-10-029-386-31985  
/ Sequence 31985, Application US/10029386  
/ Publication No. US20030194704A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Penn, Sharon G.  
/ APPLICANT: Rank, David R.  
/ APPLICANT: Hanzel, David K.  
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
/ TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
/ FILE REFERENCE: AEOMICA-X-2  
/ CURRENT APPLICATION NUMBER: US/10/029,386  
/ CURRENT FILING DATE: 2001-12-20  
/ NUMBER OF SEQ ID NOS: 34288  
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
/ SEQ ID NO 31985  
/ LENGTH: 742  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ OTHER INFORMATION: MAP TO AC010328.4  
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7  
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8  
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8  
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7  
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
/ OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALU 0.00e+00  
US-10-029-386-31985

Query Match 17.4%; Score 387.5; DB 14; Length 742;  
Best Local Similarity 30.7%; Pred. No. 8.6e-20;  
Matches 112; Conservative 46; Mismatches 154; Indels 53; Gaps 16;

QY 11 SSI-SSSSSRPKYITCTYEGCDKAVNRPSLLEQHLKRTSNDRPYKCTVDDCDKAFPRK 69  
Db 289 SSIAHQATHSGEKPYKC--NECGKVFQNSHLTNHRIHTGKPYKC--NECGKAFGVR 344  
QY 70 SHLETHIVSHSEKRPFCOSVCGKGVNSRQHLKREHITHT--KSFCTFENCOAFYKHQSS 127  
Db 345 SSIAHQATHSGEKPYKC--NECGKVFQNSHLTNHRIHTGKPYKC--NECGKAFGVR 402  
QY 128 L-BHHILSVHEKTLTCKQCNKVFTRPSTLAQHLKHHGSPAYQCDHPGCFKXFTWVSYL 186  
Db 403 LTHQVYHTGKPYKCEBGKAFNRSALYTHKRIH--TGEKPYKCEB--CGKAFNRSULTE 459

QY 187 OFH--IKOSHPRYLKCPCKCGKGVCKKGLSSHMLSHDOSTMIKIWTCDYCDVGKFAKNEL 244  
Db 460 THQVYHTGKPYKCEBGKAFNRSALYTHKRIH--TGEKPYKCEB--CGKAFNRSULTE 514  
QY 245 YNIFHDGNI PDDLKETEYVKLNNLDDQSKLNNHLETEKLVBEDEDEBEDSLD 304  
Db 515 LSRHQRIHTGKPYKCEBGKAFNRSALYTHKRIH--TGEKPYKCEB--CGKAFNRSULTE 560  
QY 305 EKRSDVSDSMASQRIKSFYASLEGSKVSGLISNGKK--INCPKNNCDMPFREVYD 364  
Db 561 -----CGKVFQNSHLAR--HRRVHTGKPYQC--NECGKAFSQTSLAR 601  
QY 365 HLKWH 369  
Db 602 HQRVH 606

Search completed: February 8, 2005, 15:34:42  
Job time : 131 secs

**This Page Blank (uspto)**





```

Db      141  CUNRRYKRNPOLRAHLNLSHNLKLTCTPHCNKSPQRYYRLRNHISKHNDEVENPVOCTAG  200
Qy      176  CFKPFQWTVLQFHFKQSHPKLCKPCGKGCVCYKGGKLSHMLSHDSDTMIKITWCDYCV  235
Db      201  CCKEERISOLQSHIKNDHPKLCPICSKPCVGENGLQMHMIHDDSLVATGNWKHICPD  260
Qy      236  GKPAKQNLVHNYNFH--DGNIPDDL-LKETEVKYLENLQGSKLNLNHLHETPKLVE  293
Db      261  MFSFKHDLTHYGSIHTEBEDPLETKYKIIDTQDL--VOHGVOLGN-----SKHSNB  312
Qy      294  EDEBEDSDLEDEKSDVRSDSMSAQRSIKSFYASLE-GSKSVSKLISGKINCPKNC  352
Db      313  QDEEKISRLRAKRRKLTENNNVPEFLQNEVDLEKRESENGNLNLLNLTVGKRRYFPYNNC  372
Qy      353  DRMSRREYDLRRHL---KMHDDNLQRIESFLNISTKEE  387
Db      373  SRTEFKTEKRYKHIDKHVYHEKLIID-----EKEE  404

```

RESULT 2  
T37676  
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37676  
R:McGouall, R.C.; Rejandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21736  
A:Accession: T37676  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-374 <MCD>  
A:Cross-references: UNIPROT:O9UTL5; EMBL:AL132675; PDB: CAB59689.1; GSPDB: GN00066; SPDB  
A:Experimental source: strain 972h-; cosmid c144  
C:Genetics:  
A:Gene: SPDB:SPAC144..09c  
A:Map position: 1  
A:Introns: 50/3  
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match	22.2%	Score 495;	DB 2;	Length 374;
Best Local Similarity	32.8%;	Pred. No. 6.7e-23;		
Matches 123;	Conservative 58;	Mismatches 160;	Indels 34;	Gaps 13;
QY	SSSSSSRRKKTICTYEGGDKAYNRP	SLLEQHLRTHSNRPYKCYVDDCDKAFPRKSHLET	74	
Db	SKNLRSAKKTICHCPREECGKKTYSR	SLLEQHLRTHSNERPVCIDYGSKAFYKRSKIKT	72	
QY	75	HIVHSEKKRPNCV--CGKGVNRQHLKRNETH--TKSEKCTFENCQAFYKQSLRH	130	
Db	73	HKRGCTNVKRPSCSHYDGCDAQFYTOQHLEHRIEVRKRPACATWEGDECFSKHQQLRS	132	
QY	131	HLISVHEKTL-----TYKQCNKLVFTRPRSKLAQHLKHNHGGSGAYOCDHGC--FKNPFQWS	184	
Db	133	HISACTHTLLRPCTYQDCBLRFATQKQLQNNVNNANHEKTIISYSPHSGVGNHGEFEKWS	192	
QY	185	VLOPHIKOSHPLKCPKCGKGVGKGLSSHMLSHDSTMIKTYCDYCDV----GKFAK	240	
Db	193	QLQNNHREAVRP-SCSICGRQFKTAAMLRHNVULH--QTLLEERTKYNCPHEGSKSPTR	249	
QY	241	KNELVEHNINHDGNIP---DDLKETEYK-KLENLDDGSKLNNHLELEFEKLVKEEDE	296	
Db	250	SSALKRHLISVHEGGMARFCHDCSGTKFGYKMLRHLRERGT-CCKAHPYINEGCIKIDG	308	
QY	297	EDEEDSLDEKSDVRSDSMS--AORSIKFTASLEGSKSVKLSINSQKINCPENNCDR	354	
Db	309	IEGAVALHQKREKELSSNLVSDVAKKLINEVTG--HGKY-----EAREYCSGSPPECNY	358	
QY	355	MFSEREYDIRRLKWM 369		
Db	359	RFKRLYDMGRHLNSH 373		

RESULT 3  
S26823  
zinc finger protein ZNF43 - human  
N:Alternate names: zinc finger protein kox27  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S26823; I37967; S10416  
R:Lovering, R.; Trowdale, J.  
Nucleic Acids Res. 19, 2921-2928, 1991  
A>Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines  
A:Reference number: S26823; MUID:91279444; PMID:1711675  
A:Accession: S26823  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-803 <LOV>  
A:Cross-references: UNIPROT:P17038; EMBL:X59244; NID:G38031; PIDN:CAA41932.1; PID:G3803332.1  
R:Thiesen, H.J.  
New Biol. 2, 363-374, 1990  
A>Title: Multiple genes encoding zinc finger domains are expressed in human T cells.  
A:Reference number: I37949; MUID:91145339; PMID:2286909  
A:Accession: I37967  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 476-531 <THI>  
A:Cross-references: EMBL:X52358; NID:G34160; PIDN:CAA6584.1; PID:G930090  
C:Genetics:  
A:Gene: GDB:ZNF43; HTP6  
A:Cross-references: GDB:128653  
A:Map position: 19p13.1-19p12  
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C:Keywords: DNA binding; zinc finger

[illegible]

RESULT 4  
G02075  
transcription repressor zinc finger protein 85 - human  
C|Species: Homo sapiens (man)  
C|Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 09-Jul-2004  
C|Accession: G02075  
R|Poncellet, D.A.  
submitted to the EMBL Data Library, September 1995

A:Reference number: G09169  
 A:Accession: G02075  
 A:Species: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residue: 1-595 <PDB>  
 A:Cross-references: UNIPROT:Q03923; EMBL:U05376; NID:g1017721; PIDN:AAA79179.1; PID:g1017721

A:Map position: 19p12-19p12  
 C:Superfamily: zinc finger protein ZFP-36, LIM metal-binding repeat homology

Query Match 17.8%; Score 396.5; DB 2; Length 595;  
 Best Local Similarity 31.7%; Pred. No. 9.9e-17;  
 Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

QY 10 ISSLSISS--SSRPKYICTYEGCDYANNRPSLLEQHLRTHSNDPRYKCTVDDCDKAFPR 68  
 DB 186 ISCLTEHSRIHTVNFYK--EECGKAFNWSSTLTGKHITGKEPKYC--EECGKAFNQ 241

QY 69 KSHLETHIVSHSKKPFHSCVCGKGVNSROHLKRNHTTT--KSFCTFENCOEAFYKIQ 126  
 DB 242 SSNLINHKHIIHGEKPKYKCECGKAFNRSTLTTHKIHTGKEPKYC--KECGKAFNRSS 299

QY 127 SL-RHHILSVHEKTLTCKOCNKVFTPRSKLAQHLKHHGSPAYQCDHPGCFNPGTWSV 185  
 DB 300 TLTHHKIHTGKEPKYKCECGKAFKOSNLTTHKIHTGKEP--YKCK--CGKAFNOSAH 356

QY 186 LQFH--IKOSHPLKCPKCGKGVGKGLSSHMLSHDSTMIKIWTCDYCDVGFPAKNE 243  
 DB 357 LTTHVEIHTGKEPKYKCECGKAFNHSHTLTHKIHTGKEPKYC--GKAFKHS 411

QY 244 LVENHVIHFDGNPDLLKTEYKLENLDDQSKLNNHELETKAYEDEDDEDESL 303  
 DB 412 TLTKHIIHTGE-----KPKSKECEKAFNOSK-----TEHKIHTGKEPYE--- 455

QY 304 DEKRSIVRSKSMASQPSIKSFYSL-----EGSK-----SVSKLISNGKIKCPKN 350  
 DB 456 CEKCGKAFNOSNLTTHKHSHTGKEPKYKCECGKAFKPSLTLTTHKIHTGKEPKYC--E 513

QY 351 NCDRMFSREYDLRHLKWH 369  
 DB 514 ECGKAFNOSKLTGKHKIH 532

RESULT 5  
 138937  
 DNA/RNA-binding protein - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
 C:Accession: I38937  
 R:Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.B.; Becker, K.G.  
 Gene 159, 215-218, 1995  
 A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA.  
 A:Reference number: I38937; MID:95347600; PMID:7622052  
 A:Accession: I38937  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residue: 1-363 <RES>  
 A:Cross-references: EMBL:U02072; NID:g644870; PIDN:AAA75623.1; PID:g644871  
 C:Superfamily: transcription factor IIA

Query Match 17.7%; Score 395.5; DB 2; Length 363;  
 Best Local Similarity 31.1%; Pred. No. 6.7e-17;  
 Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISLS-----ISSSSSSR-----PKYICTYEGCDKANNRPSLLEQHLRTSNR 53  
 DB 7 ESIVSLITADAFIAAGESSAPTPRRALPRFICSPDGSANYSKAMKDAHLCKHTGER 66

QY 54 PYKCTVDDCDKAFPRKSHLETHIVSHSEKKPFHCSV--CGKGVNSROHLKRN-EITH-- 107  
 DB 67 PFCVDEYGGCGKAFIRDVHLSRHILTHGKEPKYCAANGCDQCKNTKSNLKHFERGHENQ 126

QY 108 TKSFKTFENCOEAFYKQSLR-NHILSVHEKTLTCKQ--CNKVFTPRSKLAQHLKHHG 164  
 DB 127 QKQYICSFEDCKKTFKQHLKIHQOQHNEPLFKCTGCGGKAFNPSKLRHAAHNG 186

QY 165 GSPAYQCDHPGCFNPGTWSVLFHIIKQSHPK----- 196  
 DB 187 ----YVC-QKGSFVAKTWTTELLKHVETHEKELICEVCRKTFPKRDYLLQHMKTAPR 241

QY 197 --LKCPK--CGKGVGKGLSSHMLSHDSTMIKIWTCDYCDVGR-PAKNELVENHVI 250  
 DB 242 DVCRCRREGCGRTYTTVFNLSHILSFHEES--RPVGHAGCGKTFAMKOSLTRH-AV 297

QY 251 FHDGNPDLLKTEYK 268  
 DB 298 VHD--PDKKMKLKYK 312

RESULT 6  
 S35305  
 zinc finger protein ZNF91 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 31-Dec-1993 #sequence\_revision 02-Jun-1994 #text\_change 09-Jul-2004  
 C:Accession: S35305  
 R:Bellefroid, B.J.; Martine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; Pon-EMBO J. 12, 1363-1374, 1993  
 A:Title: Clustering organization of homologous KRAA zinc-finger genes with enhanced expr.  
 A:Reference number: S35305; MID:93223677; PMID:8467795  
 A:Molecule type: mRNA  
 A:Accession: S35305  
 A:Residue: 1-1191 <BRL>  
 A:Cross-references: UNIPROT:Q05491; EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186773  
 A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residue C:Keywords: DNA binding; zinc finger

Query Match 17.7%; Score 393.5; DB 2; Length 1191;  
 Best Local Similarity 35.8%; Pred. No. 3.2e-16;  
 Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGCDKANNRPSLLEQHLRTHSNDPRYKCTVDDCDKAFPRKSHLETHIVH 79  
 DB 905 TREKPKYC--EECGKAFSQPSHLLTHKRNHTGKEPKYC--EECGKAFSQSSTLTTHKIH 960

QY 80 SEKKPFHSCVCGKGVNSROHLKRNHTTT--KSFCTFENCOEAFYKQSL-RHHILSVH 136  
 DB 961 TGEKPKYKCECGKAFKSSLTLEHKIHTGKEPKYC--EECGKAFSQSSTLTTHKTMHNG 1018

QY 137 EKTLTCKOCNKVFTPRSKLAQHLKHHGSPAYQCDHPGCFNPGTWSVLFHIIKQSHPK 196  
 DB 1019 EKPKYKCECGKAFNRSSKLTTHKIHTGKEP--YKCE--CGKAFISSTLNGH-KIHTH 1074

QY 197 ---LKCPKCGKGVGKGLSSHMLSHDSTMIKIWTCDYCDVGFPAKNELVENHVI 253  
 DB 1075 EKPKYKCECGKAFSQSSTLTTHKRLH--TGKPKYKCGC--GKAKESSALTTHKIH 1129

QY 254 GNIPDILLKTEYKLENLDDQSKLNNHELET 287  
 DB 1130 GE-----KPKYKCECGKAFNOSSTLTTHKIH 1157

RESULT 7  
 S00647  
 finger protein - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 31-Dec-1993  
 C:Accession: S00647  
 R:Ruiz I Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.  
 EMBO J. 6, 3065-3070, 1987  
 A:Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.

A:Reference number: S00647; UID:88082679; PMID:2826129  
 A:Accession: S00647  
 A:Molecule type: mRNA  
 A:Residues: 1-1350 <RUI>  
 A:Cross-references: EMBL:X06021  
 A>Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator  
 C:Genetics:  
 A:Gene: fln  
 C:Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 375.5; DB 2; Length 1350;  
 Best Local Similarity 27.3%; Pred. No. 4,5e-15;  
 Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;

```

QY 11 SLLSSSSSRPKYCTCYEGCDKAYNRPSLLLEOHLRTSHNDRPYKTVDDCKAFRRKS 70
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 340 SFLLNQOHTSRREKPYLCSH--CNKGFIONSDLVKHRTHTGGERPYCA--ECHKGFIOKS 395
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 71 HETHIVSHSEKKRPFGCSVCGKGVNSRQHLKRNHEITHT--KSFKTFENCQEAIFYKHQSL 128
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 396 DLVKHLRTHTGGERPKFCSHCDKKFTERSALAKHQRHTGEEKPYKCS--DCGKEFTQRSNI 453
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 129 -BHHILSVHEKTLTKCKQCNKVFTRPSKLAQHLKH-----HGGSPAYOCDHPCFKN 179
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 454 ITHQRTHTGERPKYCTLCDRFTIONSDLVHQVHANLPLSDPHITANSPHKCSK--CDIT 511
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 180 FQTSVYLQPHIK--OSHPLKCPKCGKCGVCGKKGLSSHMLSH-----DD 221
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 512 FGHMSTFPMHSHKSHSEKKFQCAECCKGFTQKSDLVKHLRVHTGGERPKCLCKSPSON 571
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 222 STMIKTW-----TCYCDVGVKFAKNELVEHYNIFHDGNIPTDLLKETEVKLENI 272
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 572 SDLHKMIRHTGGERPKPYCTCD--KSFERSALIKHRT--HTGERPH-----KCSVCQKG 623
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 273 LDGSKLNNLHELETEKLVBEDEDEEDSLDEKSDVSDMSAQR-----S 320
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 624 FLOKSL-----THSRTHTGGERPKPYCTQCGKSFIONSDLVQHQRHTGGERPKYHCTEC 676
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 321 IKSFTASLEGSKSVSKLISNGKK--INCPKNCDRMFSREYDLRHLKMH--DNLQRISS 378
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 677 NKRFT---EGSSLVKKRRHSHSEKKPYRCQ--CEKFFIQSSDLVKKLVVHNGENPPAPATA 731
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 379 FLN-STEEKETPEGER 393
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 732 FHEILIRENLTRESEP 747
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 8  
 148668  
 zinc finger protein 51 - mouse  
 N:Alternate names: finger protein zfec12  
 C:Species: Mus musculus (house mouse)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
 C:Accession: I48668; A40984  
 R:Burke, P.S.; Don, J.; Wolgemuth, D.J.  
 Mamm. Genome 5, 387-389, 1994  
 A:Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of c  
 A:Reference number: I48668; UID:94319090; PMID:8043957  
 A:Accession: I48668  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Cross-references: UNIPROT:O61898; EMBL:X74855; NID:9488832; PID:CAA5247.1; PID:94888  
 A:Residues: 1-710 <RES>  
 A:Molecule type: mRNA  
 R:Crossley, P.H.; Little, P.F.R.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 7923-7927, 1991  
 A:Title: A cluster of related zinc finger protein genes is deleted in the mouse embryoni  
 A:Reference number: A40984; UID:91376058; PMID:1680234  
 A:Accession: A40984  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 672-710 <CRO>  
 A:Cross-references: GB:M74235  
 C:Genetics:

A:Gene: Zfp-51  
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
 C:Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 374.5; DB 2; Length 710;  
 Best Local Similarity 29.6%; Pred. No. 2.6e-15;  
 Matches 104; Conservative 54; Mismatches 140; Indels 53; Gaps 15;

```

QY 25 YCTEGCDKAYNRPSLLLEOHLRTSHNDRPYKTVDDCKAFRRKSHETHIVSHSEKKP 84
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 311 YEC--KEGKSPFIELSHLKRNHRTHTGGERPKYCEI--CDKSPFTTTTLTKHQKHTGKRP 366
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 85 FHGSCVCGKGVNSRQHLKRNHEITHT--KSFKTFENCQEAIFYKHQSLHLS-VHEKTLT 141
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 367 YKCRCDKSFHSHSLRHHQVHTGERPYRC--KECDKSFHBSATLRHEKHTGKTYK 424
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 142 CKQCNKVFTRPSKLAQHLKHNGSPAYOCDHPCFKNQTSVLOFH--IKQSHPKLC 199
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 425 CRBCDKSTFQRAYLKNHNRVHTGERPYECKE--CGKSFTTSTLRIHOTHTGKKPKYC 482
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 200 PKCGKGVCGKGLSSHMLSHDSTWIKITWCDYCDVGVKFAKNELVEHYNIFHDGNIPTD 259
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 483 IECGKSPFNNSYLRTHQVHSGE---KPYRCEDC-KSFTSCSTLKAHQSI-HTGEEKPYK 537
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 260 LIKETEYKLENI LDQSKLNNLHELETEKLVBEDEDEEDSLDEKSDVSDMSAQR 319
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 538 CI-----ECDKSFQVSHLRTHQVHTGERPFRCTECD-----KSFIRSHLRHQ 583
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 320 SIKSFTASLEGSKSVSKLISNGKK--INCPKNCDRMFSREYDLRHLKMH 369
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 584 KI-----HSGEKPYRC--RDCDISFSQISLNRHOKLH 614
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 9  
 A54661  
 zinc finger protein ZNF41 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 09-Jul-2004  
 C:Accession: A54661; I54227  
 R:Franze, A.; Archidiacono, N.; Rocchi, M.; Marino, M.; Grimaldi, G.  
 Genomics 9, 728-736, 1991  
 A:Title: Isolation and expression analysis of a human zinc finger gene (ZNF41) located o  
 A:Reference number: A54661; UID:91244317; PMID:2037297  
 A:Accession: A54661  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-542 <FRA>  
 A:Cross-references: UNIPROT:P51814; GB:M36548; GB:M82443; NID:9340443; PID:AAA61312.1;  
 C:Genetics:  
 A:Gene: GDB:ZNF41  
 A:Cross-references: GDB:125865; OMIM:314995  
 A:Map position: Xp11.3-Xp11.23  
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 16.7%; Score 373; DB 2; Length 542;  
 Best Local Similarity 29.0%; Pred. No. 2.4e-15;  
 Matches 115; Conservative 59; Mismatches 156; Indels 66; Gaps 21;

```

QY 14 ISSSSSRPKYCTCYEGCDKAYNRPSLLLEOHLRTSHNDRPYKTVDDCKAFRRKSHLE 73
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 163 IHQKHTBEKHNEC--NECGKAFTRKSLRNHQRHTGGERPYCA--DCGKAFIOKSHFN 218
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 74 THIVSHSEKKRPFGCSVCGKGVNSRQHLKRNHEITHT--KSFKTFENCQEAIFYKHQSLRHH 131
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 219 THQRTHTGGERPKYECDCGKSPFKSQQLVHQRHTGGERPYCT--ECGKVFHNRIN----- 272
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 132 ILVHEKTLT-----CKQCNKVFTRPSKLAQHLKHNGSPAYOCDHPCFKNQTSV 185
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 273 -LTHQKHTHTGGERPKYMACGCAFTDQSNLLKHQKHTHTGKRP--YKCN--GGCAKATWRSR 328
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 186 LQFHFKQSH---PKKCPKCGKGVCGKGLSSHMLSHDSTWIKITWCDYCDVGVK-PAKK 241
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 329 LKIH-QKSHIGRHYECKDCGKAFIOKSTLSVHORIH--TGEEKPYVPEC--GKAFIOK 382
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

QY 242 N L V E H Y N F I P H D N I P D ----- L L K E T V K L E N L D G S K L N N L H E ----- 284  
 Db 363 S H F I A H R I - H Y G E K Y E C D C G K C F T Y K S Q L - R V H Q K I H T G E K P N I C A E C C K A F T D R S N 440  
 QY 285 - L E T E K L K Y E D E D E D E D S ----- L D E K R S D V R S D S M S A Q R S I ----- K S F T A S L E G S K S 333  
 Db 441 L I T H Q K I H R E K Y E G D C G K T F T W K S R L N I H Q S H T G E R Y E C S K C G A F I Q - K A T I L S 498  
 QY 334 V S K L I S N S G K K I N C P K N N C D R M F S R E Y D L R H L K W H 369  
 Db 499 M H Q I I H T G K K P Y A C - - T E C Q K A F T D R S N L I K H Q S M H 532

RESULT 10  
 B32891  
 finger protein 2, placental - human  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Oct-1989 #sequence\_revision 20-Oct-1989 #text\_change 01-Dec-2000  
 C:Accession: B32891  
 C:Belletroid: E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A.  
 DNA 8, 377-387, 1989  
 A:Title: The human genome contains hundreds of genes coding for finger proteins of the K  
 A:Reference number: A32891; MUID:89377476; PMID:2505992  
 A:Accession: B32891  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-651 <BEL>  
 C:Cross-references: GB:M27878  
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
 I:Keywords: tandem repeat

```

Query March 16.74; Score 373; DB 2; Length 651;
Best Local Similarity 27.4%; Pred. No. 2,9e-15;
Matches 128; Conservative 57; Mismatches 146; Indels 136; Gaps 24;

QY 23 KKYICTYEGCDKAVNRPSLEQHLRTHSNDRPYKCTYDDCDKAFPRKSHLETHIVSH-- 79
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 174 KPYNCSQ--CGKAFPSQSLTSHQRTHTGKEKPYEC--GEGCKAFSRKSHLISHMRTHTGE 229
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 80 -----SRK-----KPHGCVCGKGVNSRQHLKRNHETTH--TSEFK 112
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 230 KPYGNCNCGRAFSERKSNLINHQRHTGKEKPEECGCKAFSRKSQLVTHHRHTHTGTPFG 289
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 113 CTPENCQBAFY-KHQSILRNHLISVHEKTLTKGCKNKFVTRPSKLAQHKLKHGGSP---- 167
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 290 CS--DCKRAPEKSEKELIRHQTHTGKEKPEECSEKRAKPRSSSLINHQRHTGKEKPGCT 347
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 168 -----AYGCDHPG-----CPKNEQTWSVLQFHIKQSHPKLK--CPKCG 203
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 348 QCGKAFSOKSHLISHQWHTGKEKPFCSKCGKAFSRKSQLVRH-QRHTHTGKEKPYEGCEG 406
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 204 KGCYGKGLSHTMISHDDSTMIKITWCDYCDVGK-PAKQELVHNVIIFHDGNIIDLL- 261
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 407 KAFSEKSLTNHQRIH--TGKPYVCSEC--GAFQCKSHLISHQRT-HTGKPYECSE 460
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 262 -----KETEYVKLENYLLQGSGLNLTNHELLETLEKLVDEDEDEE 300
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 461 CGKAFGEKSLATHTQRHTHTGKEKPYECRDCERAFAPQKQLNT-HQ-----RIHTGKPYE 513
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 301 DSLDEKSDVNSDSMSAQRSIKSFYASLEGSVSKLINSNGKY-INCPNNNCDRMFSRE 359
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 514 CSLCRK-----AFPEKSEKELIRHRTHTGKEKPYEC--NECKAPREK 552
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 360 YDLARHLKMHDDNIQRITESFLNLSIEKETTPGEGLVAKKARDLLPNE 406
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 553 SSLINH-----QRHTHTGKEKPEECSEC--GKAFSRKSH-LIHPQ 587
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 11
JN0533
finger protein pMLZ-4 - mouse
Species: Mus musculus (house mouse)

```

C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C/Accession: JN0533  
R/Brady, J.P.; Piatigorsky, J.  
Gene 124, 207-214, 1993  
A/Title: Cloning and characterization of a novel zinc-finger protein encoding cDNA from  
A/Reference number: JN0533; MUID:93185925; PMID:844344  
A/Accession: JN0533  
A/Molecule type: mRNA  
A/Residues: 1-393 <BBA>  
A/Cross-references: UNIPROT:Q03309; GB:M88502; NID:G200406; PID:AAA39949.1; PTD:G20040  
A/Experimental source: eye lens  
A/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C/Keywords: DNA binding; eye lens; zinc finger  
E/1-62/Region: acidic  
E/63-391/Region: zinc finger

Query Match	16.7%: Score 372; DB 2; Length 393;
Best Local Similarity	29.4%: Pred. No. 1.9e-15;
Matches 110; Conservative	57; Mismatches 163; Indels 44; Gaps 17
QY	13 LISSSSSRPKVICTYEGCDKAYNRBLLBOLHTSHSNDPRKYCTVDCDKAFPRKSHL 72
Db	49 LVHKEAAGIRYHISQ--CGKAFSQISDLIRHOKHTHGDDPRKYC--ECGSGFGRSSHLL 104
QY	73 ETHIVSHSEKKPFHCSVCGKGVNROHLKHEIHT--KSPKTPFENGQEAFLYKQSL-R 129
Db	105 IQHQRHTGKEPRYCNBEGSGFGRSSHIIQHTHTHTGEPKPKCT--EAKASAPSPHILQ 162
QY	130 HHILSVSEKTLTTCQCKNVFTRPRLKLAQHKHNGSGPAYOCDHPGCFKPNQIWS--VLQ 187
Db	163 HQHTHSEKPRYEGCEGSGFSRSHLHQHTHTGKEK-VECHN--CGRGFSERSDLLKH 219
QY	188 FHIKQSHPRKLCPKCGKGVCGKKGLSSHMLSHDSTMIKTIWTCYCDVYGKRAKKMELYEH 247
Db	220 YRVHTGPRYPKCDCEGKGNFSQNSDLVRRRAH---TGKPRPHCNVECG--ENFSRISHLVQH 275
QY	248 YNIPHDENIPDILLKETEVKKLKVILLDQGSGLNMLHELETEKLVKEBDE----EDEBDS 302
Db	276 QRT-HTBE-----KPYECTACGSGFSRSHLITQHKHTHTEKPRPECNBCKMRSPFERB 328
QY	303 LDEKRSDVRSDSMAQSRISKFTLASLGSKSVSKLIS---NSGKK-INCPKNNCDRMFS 357
Db	329 IKHQRTHTGKEPRYECVCGKGFQTS-----SNLITQHVHTGKEPRPEYEC--TECDKSF 379
QY	358 REYDLRHLKMHND 371
Db	380 RSSALIRKRAVHTD 393

```

RESULT 12
G01496
transcription factor IITA - human (fragment)
N/Alternate names: TP11A
C/Species: Homo sapiens (man)
C/Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C/Accession: G01496
R/Becker, K.G.
submitted to the EMBL Data Library, August 1994
A/Reference number: G07520
A/Accession: G01496
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-338 <BEC>
A/Cross-references: EMBL:U01434; NID:g551534; PIDN:AAA21873.1; PID:g551535
C/GeneticB:
A/Gene: GDB:GTF3A; TP11A
A/Cross-references: GDB:434744; OMIM:600860
A/Map position: 13q12.3-13q13.1
C/Superfamily: transcription factor IITA

Query Match      16.4%; Score 364.5; DB 2; Length 338;
Best Local Similarity 29.0%; Pred. No. 4.6e-15;
Matches 89; Conservative 49; Mismatches 102; Indels 67; Gaps 11

```

```

OY      KSIISL-----ISSSSSR-----PKXYICYECCDKNANRPIILBOJLATHSMDR  53
Db      8 KSIISL-----ISSSSSR-----PKXYICYECCDKNANRPIILBOJLATHSMDR  53
Db      7 ESVSLLTIADAFIAIGESSAPTPPPRPALPRFFICFPDSCANYSAMKLDMLCKHGER  66
OY      54 PYCTVDODCAFFPKSHLETHIVSHSEKDPHCSV--CGKGVNSRÖHLKEH-ETI---  107
Db      67 PFVCDYECGCAPIFRDYHLSRHIIILTHTEKEPFVCANANGCDOKFNTKSLIKGFPERKHENQ  126
OY      108 TKSFCTEENOEALFYKHQSLR-----HHILSVH  136
Db      127 QKQYICSEDECKRTYKHHQOLKIHOCCQHTNEBPLFCLOQKSCSPVAKTWTBLKHVRETH  186
OY      137 EKTLLTCQKQNVFTTPSPKLAQHKLKHGSGPAYQCDHGCFFKNPQJWSVILQFHIKQSHPK  196
Db      187 KEELICEVCRKTFPKKIDYLIKQMKHTAPERDVYVCHCPRECGCRITVTVFNLSHLSFTEE  246
OY      197 LK---CPK--CGKGCVGKKGLSSHLSHD-DSTMIKIWTCDYCDVGGFAKKNEIYEHENI  250
Db      247 SRPFVCEHAGCGKTFAMKQSLTRHAVVHDPDKKKMKL-----KYKGRERKRSILASHLS-  299
OY      251 FHDGNIP 257
Db      300 ---GYIP 303

```

```

RESULT 13
A32891
finger protein 1, placental - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 09-Jul-2004
C:Accession: A32891
R:Belletroid, E.J.; Lecocq, P.J.; Benhida, A.; Ponclet, D.A.; Belayew, A.; Martial, J.F.
DNA 8, 377-387, 1989
A>Title: The human genome contains hundreds of genes coding for finger proteins of the K
A:Reference number: A32891; MUID:89377476; PMID:2505992
A:Accession: A32891
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-428 <BL>
A:Cross-references: UNIPROT:P51522; GB:M27877
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
;Keywords: tandem repeat

```

Query Match	16.3%	Score 364	DB 2	Length 428
Best Local Similarity	29.1%	Pred. No. 6.4e-15		
Matches	116	Conservative 50	Mismatches 155	Indels 78
			Gaps 18	
QY	32	CDKAVNRSLLEQHRTHSNDRPYKCTVDDCDKAFPRKSHLETHVSHSEKKPHGCSVC	91	
Db	38	CGKLFNKKSNLASHORLHTGEKPYKC--NECGKVFHNNSHLAQHRRIHTGEKPYKCECG	95	
QY	92	KGAVNRQHLKRBHEITHT--KSPKCFENCGDAFYKHQSL--RHHLVSVEKTLTLTKQCNKV	148	
Db	96	KVFNQISHLAQHORLHTGEKPYKC--NECGKVFHQISHLAQHRITHTGEKPEYCNCKGKV	153	
QY	149	FTPRSKLAQHKLKHGGSPAYOCDDPGCFKAFQWTSVLYQFH--IKQSHPLKCDPKCGKC	206	
Db	154	FSRNSYLVQHLIHTGEKPYRCN--VCGKVFHNI SHLAQHORLHTGEKPYKCECGKVF	210	
QY	207	VGKKGSLSHMLSHDSTWIKIWTCDYCDVYG--PAKKNLVEHYNVPHQGNIP-----	257	
Db	211	SHKSLVNNHMRH--TGEKPYKCEC--GKVFHSKSLVNNHMR--HTGEKPYKCECGK	264	
QY	258	-----DDL------KETEVKKLENTLQOGSKLNNLHELETKLVEDEDEDEBDSL	303	
Db	265	VFSRNSYLAQHLIHTAGEKPYKCEBCDAFBSQNSHLVQHRITHTGEKPYKCDCEGKQFSQ	324	
QY	304	D-----EKRSQDVR-----SDMSAQRST-----KSFASLLEG	330	
Db	325	NSYLAHYHRITHTGEKAYKCEGKVFGLNSSLAHRKRIHTGEKPYKCECGKAF--SMRS	382	
QY	331	SKSVSKLISNGKINCPCNNCDRMFSREYDLRHHLKWH	369	

```

DB          363 SLTWHAIHTGEKFKC--NECGTLFRDNSYLVRHQRFH 419
               |         |||         |||         |||
finger protein (clone X1COP7.1) - African clawed frog (fragment)
S06546
C|Species: Xenopus laevis (African clawed frog)
C|Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #ext_change 09-Jul-2004
C|Accession: S06546
R|Netfield W., El-Barrati, T., Mentzel, H., Pieler, T.; Koester, M.; Poeling, A.; Knoech
J. Mol. Biol. 208, 639-659, 1989
A|Title: Second-order repeats in Xenopus laevis finger proteins.
A|Reference number: S05632; PMID:90040698; PMID:2509712
A|Accession: S06546
A|Status: not compared with conceptual translation
A|Molecule type: mRNA
A|Residues: 1-615 <NIE>
A|Cross-references: UNIPROT:P18751
C|Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
;Keywords: DNA binding; zinc finger

```

Query Match	16.3%;	Score 363.5;	DB 2;	Length 615;
Best Local Similarity	27.6%;	Pred. No. 1e-14;		
Matches 119;	Conservative 48;	Mismatches 165;	Indels 99;	Gaps 18

```

OY      1  SSLSSSSSSSKPKKTYCTYEGCDKAYNRPSSLEQHLRTHTSNRPYKCTYDDDDKAFRKS  70
Db      128  SSLVNHQRTHTGKEPFCSE--CDKCFASSSELNIHQRTHTGKPFSCS--EGCKCTNHS  184
OY      71  HLEETIVSHSEKKRPHCSVCVGKGVNRSOHLKHEITHT--KSPKCTFENCQEAIFYGQSL  128
Db      185  HPAHHQMTHTGKEPFCSCGKCGKCFASSSDLTGHRHTREKTFSCS--ECGKCFSHSHL  242
OY      129  -----RH  130
Db      243  ARHQMTHTGKEPFCSCSECGKCFSSSSGLTAHQRTHMKVPCSCAGKCFNSRSHLHH  302
OY      131  HILSVHEKTLCKOCNKVTPTRSKLAOKHLKNGSSPAVOCNPGCFKNPOTWSVLOFHI  190
Db      303  QMITGKEPFCSEFCERKCFNSPNSLARHQMTHTGKEP--FSCGE--CGKCFASSSDLTFFH  358
OY      191  KQSHPKLK--CPKCGKCGKGVK--GLSSHMLSHDDSTMIKTIQCYCDVCGPAKKNELVE  246
Db      359  HRTHTGKEPFCSECGK--CYSKSSSLVNHQRTHTGKEPFCSCSKDKC----PASSESLL--  411
OY      247  HYNIFHDQNIIDDLLKETEVEYKLELYLDQGSGLTNLHLETFEKLAVBEDBEDBDSLDK  306
Db      412  --NIHQRTHTKEAPFCSECGKCFPTNRSQLSRHQMTHTGGE--PPISCPBECECFVSS---  464
OY      307  RSDVRSDEMSAQRISIKSTFASILEGSKVS-----KLISNGCKKINCPKNNCDPMFAR  358
Db      465  -SGLTAHQOQQAHRWVKPFC--LECKCFCSNNSNFPARHQMTHTGKEPFC--SECRKRGFEN  520
OY      359  EYDLRRHLKMH  369
Db      521  QSSLARHQMTH  531

```

```

RESULT 15
A35659
knueppel-related protein H-plk - human
C:Species: Homo sapiens (man)
C:Date: 28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 09-Jul-2004
C:Accession: A35659
R:Kato, N.; Shimotohno, K.; VanLeeuwen, D.; Cohen, M.
Mol. Cell. Biol. 10, 4401-4405, 1990
A:Title: Human prostatic mRNAs down regulated in choriocarcinoma encode a zinc finger pro
A:Reference number: A35659; MUID:90318410; PMID:2115127
A:Accession: A35659
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <KAT>

```

A:Cross-references: UNIPROT:Q02313, GH:M55422, NID:9184342, PIDD:AAA36010.1, PID:g184343  
C:Superfamily: zinc finger protein ZFP-36, LIM metal-binding repeat homology

Query Match	16.1%;	Score 359.5;	DB 2;	Length 427;
-------------	--------	--------------	-------	-------------

Qy	20	RAPPKYICTYBECDCAXNRPRLLEQTLRHSNDRPYKCVADODCKAFPFKSHLTHETIVSH	79
Db	132	TRVNFYKC--EYVGFAPFNWSSLTNKRKRJHTEBKEPKVC--KEGKAFNOSTLIRKRIJH	187
Qy	80	SEKPEFHCVSCKGVNSRQHLKRHEITHTKS--FKCTFENCOEAFYKHQSL-RHNLISVH	136
Db	188	TEBEKYKCEEGCKAFNOSTLTHTHNIHTGELPYKC--EYVAFAPNOSTLIEHKLIHTG	245
Qy	137	EKTLTKOCONKVFTEPPSKLAQHKLKH-----HGSPAY	169
Db	246	EKRYACEEGCKAFNNSSKLTENKYLITGSKLYKCEEGCKAFNOSTLTHTHKRIHSEKRY	305
Qy	170	QCDHGCERKNFQTWSLQFPH--IKOSHPTLCKPCGCKGCVGKKGLSSHMLSHDSDTWIKI	227
Db	306	KCEB--CGKAFQFOSULTDHKKLIHTGEKPYKCEEGCKAFNQSJNLTLRHHVYIH---TGEXP	360
Qy	228	WTCYCDYGVKFAKKNELEVEHYNIFHDGNIIPDLLKETEVYKYLENLIDQSKANLHLEET	287
Db	361	YKCGEC--GKANOSSALNTHKRIHTGENPHKCRBEGKVFHL-----SSKLTCKKIRHT	412
Qy	288	EKL-KVEE	294
Db	413	GKLTCKEB	421

Search completed: February 8, 2005, 15:22:57  
Job time : 51 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:11:59 ; Search time 179 Seconds

(without alignments)  
1178.640 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229  
Sequence: 1 MSRSDETKSISLSISSSSSS.....PLVKARMDLLPNETVYSR 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1257	56.4	452	2	06BU9
2	850	38.1	472	2	06CK1
3	837.5	37.6	429	1	TF3A.YEAST
4	819	36.7	476	2	Q758N3
5	809.5	36.3	394	2	Q6FS08
6	495	22.2	374	1	TF3A.SCHPO
7	470.5	21.1	655	2	06C2D4
8	411.5	18.5	634	2	06R5P3
9	410.5	18.4	634	2	06PEP4
10	410.5	18.4	634	2	06BI85
11	407	18.3	611	2	Q7RY67
12	400	17.9	404	2	08N703
13	400	17.9	803	1	ZN43.HUMAN
14	398	17.9	574	1	Z492.HUMAN
15	397.5	17.8	395	2	Q436Y3
16	397.5	17.8	423	1	TF3A.HUMAN
17	396.5	17.8	595	1	ZN85.HUMAN
18	395	17.7	616	1	ZN93.HUMAN
19	393.5	17.7	1191	1	ZN91.HUMAN
20	392.5	17.6	542	2	081YV0
21	392	17.6	555	2	096N58
22	392	17.6	644	2	08ND40
23	392	17.6	661	2	08NB04
24	391	17.5	592	2	08NB04
25	390	17.5	540	2	08N211
26	390	17.5	576	2	096N22
27	389	17.5	891	2	06ZNA1
28	388.5	17.4	501	2	Q9VG72
29	388	17.4	568	2	Q9VGD3
30	387.5	17.4	839	1	Z347.HUMAN
31	385.5	17.3	297	2	08BU76

32	385	17.3	576	1	Z431.HUMAN	Q8t32	homo sapien
33	384	17.2	753	2	06P8I0	Q6P8I0	mus musculus
34	383	17.2	913	1	Z228.HUMAN	Q9UJ33	homo sapien
35	382	17.1	511	2	Q8VEE1	Q8VEE1	mus musculus
36	381	17.1	493	2	Q92IH6	Q92IH6	mus musculus
37	379.5	17.0	710	2	Q91VP4	Q91VP4	mus musculus
38	379	17.0	600	2	Q8CE00	Q8CE00	mus musculus
39	379	17.0	766	2	Q8CAJ2	Q8CAJ2	mus musculus
40	378.5	17.0	914	1	ZF62.MOUSE	Q8C827	mus musculus
41	378	17.0	393	2	Q923E2	Q923E2	mus musculus
42	378	17.0	452	2	Q8BP00	Q8BP00	mus musculus
43	378	17.0	452	2	Q8BQC2	Q8BQC2	mus musculus
44	378	17.0	461	2	Q692C6	Q692C6	mus musculus
45	378	17.0	613	2	Q9H7R5	Q9H7R5	homo sapien

## ALIGNMENTS

## RESULT 1

ID	06BU9	PRELIMINARY;	PRT;	452 AA.
AC	06BU9;			
DT	25-OCT-2004 (TrEMBLrel. 28, Created)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	Similar to sp P39933 Saccharomyces cerevisiae YPR186c TRC2			
DE	TFIIIA.			
GN	ORFNames=DEHA0C07623g;			
OS	Debaromyces hanseni CBS767.			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Debaromyces.			
OX	NCBI_TaxID=284592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CBST67;			
RG	Genolevures;			
RA	Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marcq C., Neveglisse C., Talia E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckertich J.M., Beyne E., Bleykasten C., Boistrame A., Boyer J., Cattolico L., Confanioleri F., de Darvar A., Despons L., Fabre B., Fairhead C., Ferry-Dumazet H., Groppi A., Hantaye F., Hennequin C., Janniaux N., Joyet P., Kachouri R., Karrest A., Kozul R., Lemaitre M., Lesur I., Ma L., Muller H., Nicaud J.M., Nikolski M., Ozdas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swemene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zeniener-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Galliard C., Weisenbach J., Winkler P., Souciet J.L., "Genome evolution in yeasts."			
RL	Nature 430:35-44(2004).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CBST67;			
RA	Genoscope;			
RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; CR82135; CAG86038.1; -.			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0003676; F:nucleic acid binding; IEA.			
DR	GO; GO:0008270; F:zinc ion binding; IEA.			
DR	InterPro; IPR007087; Znf_C2H2.			
DR	Pfam; PF00096; Zf_C2H2_8.			
DR	SMART; SM00355; Znf_C2H2_9.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.			
DR	PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 7.			
SO	SEQUENCE 452 AA; 52049 MW; D8224533C04B0269 CRC64;			

Query Match 56.4%; Score 1257; DB 2; Length 452;  
Best Local Similarity 55.6%; Pred. No. 2,5e-64;  
Matches 234; Conservative 62; Mismatches 93; Indels 32; Gaps 10;



```

QY 3 ESDETKSISSLSSSSSRPKKYICTYEGCDKAYNRPSSLEQHLKTHSNDRPYKCTVDDC 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 ESEETIASVSR-ASSSSARPKNYICBPDGKAYSPSLBQKRRHTRRPKCSBTGC 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 DKAFPRKSHLETIVSH--SEKKPFQSVCGKGVNSRQHLKREHITHTYSFKCTFENCOS 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DKSFLLKSHLQAHLLSHEDESKPFQCATGKGVNTLQHLKREHITHTYSFICTFEGCN 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 APTKHOSLRHHILSVHEKTLTKQCKNVFTRPSKLAQHLKHHGSDPAYQCDHPGCFKPF 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 SFYKHOSLRHHITLSVHEKTLCKNKCSFNRPYRLAQHNIKYHSDSPAYQCDHQGCFGN 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 QMTSVLOPHIKOSHPRKLCCKPGCKGVCKKGLSSHMLSHDSTMTIKMTCDYCDYKFAK 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 MTMSALQHLKTHPRKICPCTCGKGVCKKGLSHNNHDEEKIVKLAMCNYNICKFPYK 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 KNELVHYNIPHDGNIPDDLKETEYKLENLIDQGSK----LNNHLETERK-LKVEED 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 KADLIHNTYTHDRNLPDDLKPIREQLDKLSBKDSTNIDPNTLEGLOSKGFVPSD 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 296 EDEEDSLDEKRSVDVSDMSA-QRSIKSFYASLE-GSKSVSKLISNS--GKKINCPKON 351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 EEKKE-----NDYMSNGLOASHKSLNSLNTLESGKASIVDLLNNYLLKRVPCPKON 353
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 352 CDMFERREYDLRHLKMHNDNLORISPLANSIKEET-----PEG-EPVYKKA 398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 CDMFERDYDLERHLKMHESHLKKIEAFKSLLEBKVPDSNMHDKRLPEPDEPYMAWK 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 399 R 399
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 R 414
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 2

```

OCWK1 PRELIMINARY; PRT; 472 AA.
ID 06CWL;
AC 06CWL;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Similarities with sp|P39933 Saccharomyces cerevisiae YPR166C TFC2
DE TFI1A.
GN ORFNames=KLIAB03454g;
OS Kluyveromyces lactis NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBITaxID=284590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marcq C., Neugegola S.,
RA Coffard N., Frangeul U., Aigle M., Anthouard V., Babout A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer J., Catolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrast A., Koszul R., Lemaire M., Leeur I., Ma L., Muller H.,
RA Nicard J.M., Nikolski M., Oztas S., Olier-Kalogeropoulos O.,
RA Pellenn S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenneu D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenoune-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardrin C., Weisenbach J.,
RA Wincker P., Souciet J.L.
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope; (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, CR382122, CAH02081.1, -

```

```

DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR SMART; SM00355; ZNF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
SQ SEQUENCE 472 AA, 54229 MW, 5AB712DB2674A4D CRC64;

```

Query Match 38.1%; Score 850; DB 2; Length 472;  
 Best local similarity 45.3%; Pred. No. 5.6e-41;  
 Matches 169; Conservative 48; Mismatches 128; Indels 28; Gaps 7;

```

QY 9 SLSLSISSSSSSRPKKYICTYEGCDKAYNRPSSLEQHLKTHSNDRPYKCTVDDCDAKF 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 90 STSTSVSVSSSQRSKYVFCDDYQCKNSFTFRPBLTLEHQLTVHNGIKPFC--DTGKERA 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 68 RKSHEETHIVSHSEKKPFQSVCGKGVNSRQHLKREHITHTYSFKCTFENCDAFVQGS 127
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 KXSHLRHMFSTTDKPFQCSICGKVYTRQOLKREHITHTYSFQSVYGCNESPYPKHPQ 207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 128 LRHHILSVHEKTLTKQCKNVFTRPSKLAQHLKHHG--GSPAYQCDHPGCFKPFQTSV 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 LRSHILSVHEKTLTCPCNKCTFORPYRLKNHIDKHNPESTGMVQCDFLSTDVFSWSS 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 186 LQPHIKOSHPRKLCCKPGCKGVCKKGLSSHMLSHDSTMTIKMTCDYCDYKFAKKNELV 245
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 LQPHIKQCHPKLPCTICGKPCVASEGLNHHMHHBESLVTKMKKSSCPDTSFAKKTQV 327
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 246 EHYNIPHDGNIPDDLKETEYKLENLIDQGSKLNHLETERKLVBEDEE----- 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 QHYESHKVDVPELLELLPTDADP-----DTNKKVNT-----DKISTEBSFSGKKT 375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 299 --EEDSLDEKRSVDVSDMSAQRSIKSFYASLESGKSVSKLISNSGKKINCPKONCDPMF 356
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 376 LIDPPSPSKRVDELGSVEAEIKIKYTES--GKTVSLNLTWQGRKCPYTNCKRFT 433
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 357 SREYDLRRLKWH 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 434 KSKKEFDTHIEKH 446
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 3

```

TF3A_YEAST STANDARD; PRT; 429 AA.
ID TF3A_YEAST
AC P39933;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 05-JUL-2004 (rel. 44, Last annotation update)
DE Transcription factor TFI1A (TFI1A).
GN Name=TFI1A; Synonyms=P2FL, TFI1A; OrderedLocustNames=YPR166C;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBITaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92147684; PubMed=1737784;
RA Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA Segall J.E.
RT "The deduced sequence of the transcription factor TFI1A from
RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT TFI1A."
RL J. Biol. Chem. 267:3282-3288(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=92237295; PubMed=1570325;
RA Woychik N.A., Young R.A.;
RT "Genes encoding transcription factor TFI1A and the RNA polymerase
RT common subunit RPB6 are divergently transcribed in Saccharomyces
RT cerevisiae."

```

RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003 (1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=8288C / AB972;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,  
RA Azaou R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duestenhoef A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hepling U., Heumann K., Hilbert H., Hillier L.W.,  
RA Hunkeler-Smith S., Hyman R., Johnston M., Kalkan S., Kleine K.,  
RA Komp C., Kundi O., Laebkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marabe R., Messenguy F., Mewes H.-W., Mitrinac S., Moestl D.,  
RA Mueller-Auer S., Namath A., Netewich U., Oefner P., Pearson D.,  
RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,  
RA Scherens B., Schramm S., Schroeder M., Sidcu A.M., Tettein H.,  
RA Uregetarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambut R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.W., Zollner A., Vo D.H., Hani J.,  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.",  
RL Nature 387:103-105 (1997).  
CC -I- FUNCTION: Interacts with the internal control region (ICR) of  
CC approximately 50 bases within the 5S RNA genes, is required for  
CC correct transcription of these genes by RNA polymerase III. Also  
CC binds the transcribed 5S RNA's.  
CC -I- SUBCELLULAR LOCATION: Nuclear.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M80611; AAB08014.1; -;  
DR EMBL: M90638; -; NOT ANNOTATED\_CDS.  
DR EMBL: U25841; AAB64615.1; -;  
DR PIR: S20050; S20050.  
DR HSSP: P25490; IUBD.  
DR Germonline: 144451; -;  
DR TRANSFAC: T03530; -;  
DR SGD: S000006390; PZFL.  
DR GO: GO:0005677; C:transcription factor complex; IPI.  
DR GO: GO:0003709; F:RNA polymerase III transcription factor act. .; IPI.  
DR GO: GO:0006384; P:transcription initiation from Pol III promoter; IPI.  
DR InterPro: IPR007087; Znf C2H2.  
DR Pfam: PF00095; zfc-C2H2; 5.  
DR SMART: SM00355; Znf\_C2H2; 9.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 6.  
DR DNA-binding; Metal-binding; Nuclear protein; Repeat; RNA-binding;  
KW Transcription regulation; Zinc-finger.  
FT DOMAIN 24 43 Ser-rich.  
FT ZN\_FING 49 74 C2H2-type.  
FT ZN\_FING 80 102 C2H2-type.  
FT ZN\_FING 108 130 C2H2-type.  
FT ZN\_FING 134 159 C2H2-type.  
FT ZN\_FING 163 186 C2H2-type.  
FT ZN\_FING 194 219 C2H2-type.  
FT ZN\_FING 222 244 C2H2-type.  
FT ZN\_FING 253 277 C2H2-type.  
FT ZN\_FING 321 327 Arg/Lys-rich (basic).  
FT DOMAIN 365 389 C2H2-type.  
FT ZN\_FING 429 50027 MW; 209B1EDBA20422D9 CRC64;  
SQ SEQUENCE 429 AA; 50027 MW; 209B1EDBA20422D9 CRC64;  
Query Match 37.6%; Score 837.5; DB 1; Length 429;  
Best Local Similarity 43.7%; Pred. No. 2; Ce-40;  
Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;  
QY 1 MSBDETKSISLIS--SSSSSRPKYICTYEGCDKAYNRPISLLEQH-LATHSNDPRPYKC 57

Db 23 ISRSESESINSLTSTRSSSSNRPKTYPCDYDGCDAFTRPISLTHQLSVHGLAFAQC 82  
QY 58 TYDDCKAFPRKSHLETHIVSHSEKPRFHCSCGVKAVNRQHLKREIHTHTSFKCTFEN 117  
Db 83 --DKCKASFVKSHLERHLVTHSDTRPFQCSYCGKAVTTRQDLKRHEVHTHTSFIPEEG 140  
QY 118 CQEAFFKQSLRNHLISVYEKTLCTCKQCNKVFTRPEKLAQHLKHGG--SPAYQCDHPG 175  
Db 141 CNLRFYKHQQLRAHLISVHLAKTLCTCPNCKSFQRPRLRNHLSKHHDEVENYQCTFAG 200  
QY 176 CFKQFQWTVLQPHIKQSHPKLCKPCGKGVCKGLSASHMLSHDSTWIKITWTCYCDV 235  
Db 201 CCKEPFHWQLQSHIKNDHPKLCFICSPYCGENGLQNMHIIHDSLVTKWKKHICPD 260  
QY 236 GKPAKNELVENYNIH-DGNIPDD-LKTEYKLENLIDQSKNLNHELETUKAYE 293  
Db 261 MFSFRKHDLTHYGSJTHEEDIPLKLYKISDIOO--VQDHGVQIGN-----SKHSNE 312  
QY 294 EDEDEDESDLEKRSQVRSQMSAORSIKSFASLE-GSKYSKLSNSGKKINCPKNC 352  
Db 313 QDEKISNLRKRRKRLTENNVEFLQNEVDLEKRLSEGENGLNLNTVGRKTRCYNNC 372  
QY 353 DRMFREYDLRRHL--KWHDDNLQRISEFLNSIEKEE 387  
Db 373 SRTFKTKKYEKXKIDGKHVELKTLQ-----EKEE 404  
RESULT 4  
Q758N3 PRELIMINARY; PRT; 476 AA.  
ID Q758N3;  
AC Q758N3;  
DT 05-JUL-2004 (TReMBLrel. 27, Created)  
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)  
DE AEL278WP.  
GN ORFNames=AEL278W;  
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.  
OX NCBI\_TaxID=33169;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 10895;  
RA Gates K., Dietrich F.S., Brachat S., Voegeli S.E., Lerch A.,  
RA Phillipsen P., Gaffney T.;  
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE016818; AAS52406.1; -;  
DR HSSP: P03001; IUBG.  
DR AGD: AEL278W; -;  
DR GO: GO:0005634; C:nucleus; IEA.  
DR GO: GO:0003767; F:nucleic acid binding; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR InterPro: IPR007087; Znf C2H2.  
DR Pfam: PF00095; zfc-C2H2; 8.  
DR SMART: SM00355; Znf\_C2H2; 9.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 3.  
DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 3.  
DR DNA-binding; Metal-binding; Nuclear protein; Repeat; RNA-binding;  
KW Transcription regulation; Zinc-finger.  
FT DOMAIN 476 53433 MW; D1467545D4E6BE0D CRC64;  
SQ SEQUENCE 476 AA; 53433 MW; D1467545D4E6BE0D CRC64;  
Query Match 36.7%; Score 819; DB 2; Length 476;  
Best Local Similarity 40.8%; Pred. No. 3; Ae-19;  
Matches 165; Conservative 65; Mismatches 136; Indels 38; Gaps 9;  
QY 2 MSBDETKSISLISSSSSSRPKYICTYEGCDKAYNRPISLLEQHLT-HSNDPRPYCTVD 60  
Db 96 TSAGSTVSLSSVSTASTSTRKTHVCDYEGCYKAFTRPISLTHQGTAGGRAQGC--E 153  
QY 61 DCDKAFPRKSHLETHIVSHSEKPRFHCSCGVKAVNRQHLKREIHTHTSFKCTFENCOE 120  
Db 154 QCGRGFTKSHLERHLFHSSEKTPFSCVCGKAVTTRQDLKRHEIHTHTSFKCPHGGCB 213  
QY 121 AFYKHQSLRNHLISVYEKTLCTCKQCNKVFTRPEKLAQHLKHGGSPA-----YQCDHPG 176

Db 214 AFKPHQLSHVLAHVEQKLTCTHCQKRFQRPRLTTHAKHHG--PAQFQFQCTNAGC 271  
QY 177 FKNFQTSVLOPHIKOSHPELKCPCKGCVCVGGKGLSHMLSHDSTMIKIWTCDYCDVG 236  
Db 272 VQCFEETWSALQOQLHTDHPRLPGCVGCKLCVGETGQMVTYHDESRVLIKMKCSVCSOT 331  
QY 237 KFAKKEIVLVEHNYIFHDGNI PDDLKETEYKLENLIDGSKLNNLHELETEKLYEBDE 296  
Db 332 TYAKMADLLAHYKTKGDSIPKELIETHVNA-----SPAVHYVDEQQ 373  
QY 297 EDESDLDKERSDVRD-----SMSAQRISKFTASLEGSKSVKLSNGKKNCPKN 351  
Db 374 REDPELVPRQKRKNSDNTIMGSLQTEEKIRKIES--GRGTSLLLNTAGRKRCPYLG 431  
QY 352 CDRMFGR--YDLRRLKWHDDNLQRIESFLNSIEKEETPEGP 393  
Db 432 CSRVFTEDEKYL--HISKHINDLKVLLKEDLTKTEAAGNDP 473

## RESULT 5

Q6FSJ8 PRELIMINARY; PRT; 394 AA.

AC 06FSJ8; 05-JUL-2004 (T-EMBLrel. 27, Created)  
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
DE Similar to sp|P39933 Saccharomyces cerevisiae YPR18c TFC2.  
GN ORFNames=CAGL0610021g;  
OS Candida glabrata CBS138.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=284593;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBS138;  
RG Genolevures;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neugeglise C., Talla E.,  
RA Goffard N., Frangoul L., Aigle M., Anthonard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.M., Bayne E., Bleykasten C.,  
RA Boistrame A., Boyer J., Catolico L., Confalonieri F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet F., Groppi A.,  
RA Hantaye F., Hennequin C., Jaunhaux N., Joyet P., Kachouri R.,  
RA Kerest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.M., Nkolosi M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
RA Swemene D., Tekala F., Wesolowski-Louvel M., Westhof B., Wirth B.,  
RA Zenlou-Meyer W., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Galliardin C., Weissenbach J.,  
RA Wincker P., Souciet J.L.;  
RT "Genome evolution in Yeasts."  
RL Nature 430:35-44(2004)  
DR EMBL; CR380953; CAG59723.1; -;  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PR00096; zf-C2H2; 8.  
DR SMART; SM00355; Znf\_C2H2; 9.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 4.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 4.  
SQ SEQUENCE 394 AA; 45716 MW; 130B610E5FCF799 CRC64;

Query Match 36.3%; Score 809.5; DB 2; Length 394;  
Best Local Similarity 39.8%; Pred. No. 9.7e-39;  
Matches 162; Conservative 66; Mismatches 134; Indels 45; Gaps 8;

QY 1 MEESDTKSISLSSSSSRPKYICTEGCDKAVNRSLLEQHRT-HSNDRPYKCTV 59  
Db 13 LBRSSSVGMS--VAGGSVRSKTYLCEDNCDKATRSLSLTHONTYHLGKPKKC-- 69  
QY 60 DDCKAFPRKSHLETHIVSHSEKKPFCVCGVNSRQHLKRIETHTKSPKCTPENQ 119

Db 70 NCCESSFYKRIHLEHLYTHDERPFCYSCFCKGLITRQOLRRHVTHTKSNCEYEGN 129  
QY 120 EAFYGHOSLRHHILSVHEKTLTCKQCNKVFPPSKLAQHKLGHHGS--PAYQCHPGCF 177  
Db 130 EEPYGHPLRAHILVAHLOSLKCHECNKSFORPYLKHIAHNPDDVYNAVQCTFVSYS 189  
QY 178 KNPQTSVLOPHIKOSHPELKCPCKGCVCVGGKGLSHMLSHDSTMIKIWTCDYCDVG 237  
Db 190 KSPFTWSALRLHVKNDHPLKCPICSCVGEEDGIMNHKIIDENLVSRNWKHICNDQS 249  
QY 238 FAKKEIVLVEHNYIFHDGNI PDDLKETEYKLENLIDGSKLNNLHELETEKLYEBDE 297  
Db 250 FAKKEIVLVEHNYIFHDGNI PDDLKETEYKLENLIDGSKLNNLHELETEKLYEBDE 293  
QY 298 DEEDSLDKERSDVRSDSQAQRISKFTASLEGSKSVKLSNS--GKKNCPKNCDRWF 356  
Db 294 -----SPKKYKVAITDMLAITETNNLRFPGGDMATLLNTYGRKFRCCYSKCYRSF 347

## RESULT 6

TF3A\_SCHPO STANDARD; PRT; 374 AA.

AC 09UTL5; 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Transcription factor IIRA (Factor A) (TFIIRA).  
GN Name=sfc2; ORFNames=SPAC144.09c;  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RC MEDLINE=22082319; PubMed=12087160; DOI=10.1093/nar/gk385;  
RX Schuman D.B., Seizer D.R.;  
RT "Identification and characterization of transcription factor IIRA from  
Schizosaccharomyces pombe".  
RL Nucleic Acids Res. 30:2772-2781(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsle K.,  
RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch B.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymowicz B.,  
RA Wellfens I., Vanterre E., Rieger M., Schaefer M., Wellner-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer B., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler W., Wambut L., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliardin C., Tallada V.A., Garçon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Batteil B.G., Nurell B.G., Nurell B.G.,

Query Match	Best Local Similarity	Matches	Conservative	Score	DB 1	Length	DB 2	Indels	Gaps
15	SSSSSSRKKYICTYEGCDKAYNRPSSLLQHLRTHSNDPRPYKCTVDDCKAFPRKSHLET	123	58	495	DB 1	374	34	13	
13	SKNLRSAKKIFPHCPYECGCKYSRPSLLEQHLRTHSNRPVCDYTGSKAFYRKSHLKI	123	58	495	DB 1	374	34	13	
75	HIVHSEKKPCHCS--CGKVNRSQHLKRNHITL--TKSKCTPENCQAFYKQSLRH	123	58	495	DB 1	374	34	13	
73	HKRCITNVKPPSCHYDGDCAOFPYTOOHLERHIEVIRKKRPACWEGDCBEFSKQOURLS	123	58	495	DB 1	374	34	13	
131	HILSVHEKTL-----TKRCNVVFTPRPSLLQHLKHNHGSPAYOCDHDPG--FKNFWMS	123	58	495	DB 1	374	34	13	
133	HISACHTLLLPYCTYDCEIRFATKQKQNNVNAHEKIIISYSCPHSSCYGHBEFKMS	123	58	495	DB 1	374	34	13	
185	VLOPHIKOSHFKLKCPRCKGCGVGGKGSNHLSHDSTMIKTWICYDV----GKPAK	123	58	495	DB 1	374	34	13	
193	QLQNHIRAAVP--SGSICGRQFKPAALHRRHVVLH--QTLLEKRTYCHPNEGCKKSPTR	123	58	495	DB 1	374	34	13	
241	KNEIYEHYNIYHDGNIP--DDLKETEYK--KLENLDQSGKANNHLETFEKLKVEEDE	123	58	495	DB 1	374	34	13	
250	SSALKKHISVHEGMAFHCHSCGCTKPFYKMLQPHLRGT--CKKAHNPYINEGCIKIDG	123	58	495	DB 1	374	34	13	
297	EDEBDSLEKRSVDYSDMS--AQSISLFTASLEGSGSVKLTISNGKLTINCPANNCDR	123	58	495	DB 1	374	34	13	
309	IEGVAIHQKEKEKESLNSVDVAKILNEVTG--HGK-----EARYSCSPFBCNY	123	58	495	DB 1	374	34	13	
355	MFSREYDLRRHLKWH	123	58	495	DB 1	374	34	13	
359	RFKRLYDHRRHLNSH	123	58	495	DB 1	374	34	13	

ID	Q6C2U4	PRELIMINARY;	PR:	655 AA.
AC	Q6C2U4			
DT	25-OCT-2004 (TREMBLrel, 28, Created)			
DT	25-OCT-2004 (TREMBLrel, 28, Last sequence update)			
DT	25-OCT-2004 (TREMBLrel, 28, Last annotation update)			
DE	Yarrowia lipolytica chromosome F of strain CLIB99 of Yarrowia			
DE	lipolytica.			
GN	ORFNames=YALI0F05104g;			
OS	Yarrowia lipolytica CLIB99.			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Dipodascaceae; Yarrowia.			
OX	NCBI_TaxID=284591;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CLIB99;			
RG	Genolevures;			
RA	Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,			
RA	LaFontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,			
RA	Goffard N., Frangeul L., Aigle M., Anthouard V., Babor A., Barbe V.,			
RA	Barney S., Blanchin S., Beckert J.M., Beyne E., Bleykasten C.,			
RA	Bolsarine A., Boyer J., Catolico L., Confaiolieri F., de Daruvar A.,			
RA	Despons L., Fabre E., Fairhead C., Ferry-Dumas H., Groppi A.,			
RA	Hartveit F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,			
RA	Kerrest A., Koszul R., Lemaire M., Lesut I., Ma L., Muller H.,			
RA	Nicoud J.M., Nikolekt M., Oztes S., Ozler-Kalogeropoulos O.,			
RA	Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A.,			
RA	Svenneberg D., Tekaia F., Wesolowski-Jouvet M., Wesshof E., Wirth B.,			
RA	Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,			
RA	Boucher C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,			
RT	Wincker P., Soucieu J.L.,			
RL	"Genome evolution in yeasts."			
RL	Nature 430:35-44(2004).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CLIB99;			
RA	Genoscope;			
RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: CR382132; CAC77825.1; -			
DR	GO: GO:0005634; C:nucleus; IEA.			
DR	GO: GO:0003676; F:nucleic acid binding; IEA.			
DR	GO: GO:0008270; F:zinc ion binding; IEA.			
DR	InterPro: IPR006025; Pept_M_zn_BS.			
DR	InterPro: IPR007087; ZnF_C2H2.			
DR	Pfam: PF00096; zf-C2H2; 5.			
DR	Prodom: PD000003; ZnF_C2H2; 1.			
DR	SMART: SM00355; ZnF_C2H2; 10.			
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.			
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.			
DR	PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN 1.			
DR	SEQUENCE 655 AA; 74609 MW; 5FF764F312CF4882 CRC64;			
SR				
Query Match	21.1%;	Score 470.5;	DB 2;	Length 655;
Best Local Similarity	28.9%;	Pred. No. 4.6e-19;		
Matches 111;	Conservative 65;	Mismatches 145;	Indels 63;	Gaps 11;
QY	21 RPKYKICGECGDAAYNRPSSLEQHLKRNHSDREYKCTVDDCDCAFPRKSHLETIVSHS 80			
DB	83 RTLAHVCPREBCDRAISRPSLLKQHLNHSYMERFCVCTYECGCGGFRFRSHLKANTSHST 142			
QY	81 EKKDFHSGVCGKGVNRSROHLKRHEITVTKSFCTFENCQEAFFYKQSLRNHILSVH--E 137			
DB	143 VAKYHVSFCFAKGFNTRQHLDRHREVTIKSIGCHCPHODCSKQCRTEVQLLEHCVADHATWK 202			
QY	138 KTLVCKQCNKVFTRPSKLAQKLGKHGGSAPAYCDDHGCFCFNPTWSTVLQCHIQSHPKL 197			
DB	203 HQCTYPREGRAIGSERQWQYMWEDNHQTPQYKCPHDECEVASQWIKLQGVSAKHNRW 262			
QY	198 KCPKCGKGVCGKGL--SSHMLSHDST-----MIKITC--DYCDVG 236			
DB	263 LGHHC-----EGLFDDVQSMYHNSKSVHGLSTVLVYGGHRTPEGVNQMLWCPRESCHQ 315			
QY	237 KPAKKNELVEHYNI FHDGNIPD--DLKETEYV-----KLENLDDQSGSKLNNLHELE 286			

Db	Seq	Similarity	Score	DB 2	Length	634
Db	316 SF-DDDLAKHVVSHNKAPVROMFESVVAEYDITPELIIIPFEKPEKLTIRG	18.5%	411.5	DB 2	Length	634
Qy	287 TEKLVEDEDEDESDLEKRSVDSDNSAQRISKFTASLEGSKVSLISNSG	32.8%	1.1e-15			
Db	366 SERTLEGSGSEEEYOLPESR	54	Mismatches	129	Indels	57
Qy	343 KKINCPKNNCDMPSEBYDLRPHL	54	Gaps	19		
Db	418 RTTPCVETCMHFRNRYDLDRHV					
RESULT 8						
Q6R5P3	PRELIMINARY	PRT	634	AA		
ID	Q6R5P3					
AC	Q6R5P3					
DT	05-JUL-2004 (TREMBLrel. 27, Created)					
DT	05-JUL-2004 (TREMBLrel. 27, last sequence update)					
DT	05-JUL-2004 (TREMBLrel. 27, last annotation update)					
DE	KRAB-zinc finger protein.					
GN	Name=ce8					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	STRAIN=129S6/5VEVATC;					
RC	Brathwaite M., Maelitz P., Dudekula D., Qian Y., Nagaraja R.,					
RL	Submitted (DEC-2003) to the EMBL/GenBank/DBD databases.					
RL	EMBL; AY510701; AAR91692.1; --					
DR	HSSP; P08046; 1A1G.					
DR	GO; GO:0005634; C:nucleus; IEA.					
DR	GO; GO:0003676; F:nucleic acid binding; IEA.					
DR	GO; GO:0008270; F:zinc ion binding; IEA.					
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.					
DR	InterPro; IPR001909; KRAB.					
DR	InterPro; IPR007087; ZnF_C2H2.					
DR	InterPro; IPR007086; ZnF_C2H2_sub.					
DR	Pfam; PF01352; KRAB; 1.					
DR	Pfam; PF00096; zf-C2H2; 16.					
DR	PRINTS; PR00048; ZINC_FINGER.					
DR	PRODOM; PP000003; ZnF_C2H2; 15.					
DR	SMART; SM00349; KRAB; 1.					
DR	SMART; SM00355; ZnF_C2H2; 16.					
DR	PROSITE; PS00805; KRAB; 1.					
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.					
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 16.					
SO	SEQUENCE 634 AA, 73724 MW, CBB21DDE42505D3D CRC64;					
Query Match						
Beat Local Similarity 32.8%; Pred. No. 1.1e-15;						
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19						
Qy	23 KKIYCTYGCDKAINRPSLLEQHLTHSNDPRYKCTVDDCYAFPRKSHLETHIVSHSK					82
Db	323 KPYKC--KECGAFYQSSSLKGLHLHKGKKPKYC--EECDKSFTEKSTLTTHKRIHTGK					378
Qy	83 KPFHSGVSGKGNVSNQHLKRHEITHT--KSPKCTENQCEAPYKQSLR--HILISVHEKT					139
Db	379 KPYKNITGKGFNSCTNLKTHQTLHTEGPKYC--KECGKAPPYMSLTIHONLHPGEKP					436
Qy	140 LTKQCNKVFYTPSPKLAQHLGHGSGSPAYQCDHGCFFNPFQTSVLQFN--IKOSHPL					197
Db	437 YKCEGCDSPFLKSLRTHQIHTHEKRP--YKCNIT--CGKSPQCTMLKTHQHLHGEKPY					493
Qy	198 KCPKCGKCGVGGKGLSHMLSHDSDTMKIWTCDYCDVGEKFAKKNELVEYNIIFHDGNIP					257
Db	494 KCKEGRGKSFHWYSLSLKHQNLHSGE--KPYCKECD--KSPFEKSTLIGQRI--HTG---					545
Qy	258 DDLKETEYKLE--NLDDQSKLNNHLELEFEKLVGEDEDEDESDLEKRSVDSDNS					315
Db	546 -----KLLYKNCNICDK--SFTWCASLKTHK--KFNHGEK-----PY					577

Oy	316	SACSISF--TALESKSVSKXSIXSCKXI-NCPKNXCRMPRSREYDLRHLTWMH	369
Dd	578	KCRCKGKSPPOLSTLTKHQZHEHKXHTGEKLYCC-NDCCRSTSYNSSFRRHXKH	632
RESULT 9			
ID	06PEP4	PRELIMINARY;	PRT; 634 AA.
AC	06PEP4		
DT	05-JUL-2004	(TREMBLrel. 27, Created)	
DT	05-JUL-2004	(TREMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)	
DE	Hyothetical protein A830058L05.		
GN	Name=A830058L05:		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCB1_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRPAIN-FVB/N; TISSUE=Mammary tumor. C3;		
RX	MEBLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strauberg R.L., Felsing E.A., Grouse L.H., Derge J.G.,		
RA	Klauser R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,		
RA	Alesch S.F., Zieberg B., Butow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hatel F.,		
RA	Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,		
RA	Brownstein W.J., Uedin T.B., Toshitsuki S., Carrinetti P., Prange C.,		
RA	Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bohak S.A., McQueen P.T., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,		
RA	Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Faley J., Helton E., Kettman W., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywnicki M.I., Skaleka U., Smallue D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RL	and mouse cDNA sequences."		
RN	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRPAIN-FVB/N; TISSUE=Mammary tumor. C3;		
RA	Strauberg R.;		
RL	Submitted (SEP-2003) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; BC057947; AAHS7947.1; .		
DR	HSSP; P08046; IAIQ.		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003476; F:nucleic acid binding; IEA.		
DR	GO; GO:0008270; F;zinc ion binding; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPRO01909; KRAB.		
DR	InterPro; IPRO07087; znf_C2H2.		
DR	InterPro; IPRO07086; znf_C2H2_sub.		
DR	pfam; PF00352; KRAB_1		
DR	pfam; PF00096; zf-C2H2; 16.		
DR	PRINTS; PR00048; ZINC_FINGER.		
DR	PRODOM; PD000003; znf_C2H2; 15.		
DR	SMART; SM00349; KRAB; 1.		
DR	SMART; SM00355; znf_C2H2; 16.		
DR	PROSITE; PS50805; KRAB; 1.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.		
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.		
DR	HYPOTHEICAL PROTEIN.		
SO	SEQUENCE 634 AA; 73682 MW; EG7E83B452DEF24 CRC64;		
Query Match	18.4%;	Score 410.5;	DB 2; Length 634;
Best Local Similarity	32.8%;	Pred. No. 1.2e-15;	
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 191;			

QY 23 KKYCTYEGCDKAYNRPSSLEOHRLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82  
 DB 323 KPYKC--KECGKAFQTSSSLKHLNLTHTGKKPKC--EECDKSFTEKSTLTTHKRJHTGK 378  
 QY 83 KPRHCGVCGKGVNSRQHLKRHEITHT--KSFCTFENCCEAFYKHOSLR--HHILSVHEKT 139  
 DB 379 KPRKCNICGSPFNSCTNLKTHQTLHTGCKPYKC--KECGKAFYMSLSLKHQMLHNGEKP 436  
 QY 140 LTKGCKNVFTRPSKLAQHLKHHGSPAYOCDFPCFKNFQWTVLOFH--IKOSHPLK 197  
 DB 437 YKCECKDSFTLKSLSLTHQTLHTGCKP--YKCN1--CGKSFNCTNLKTHQTLHTGCKP 493  
 QY 198 KCPKCGKGVCGKGLSSHMLSHDSTWIKITWCDYCVGKFAKKNELVEHYNIFFHGNIP 257  
 DB 494 KCECKGKSFHYMSLSKSHQNLHSGE---KPYCKECD--KSFTKSTLIKQRI--HTG--- 545  
 QY 258 DDLKETEYKLE--NLDDGSKLNNLHELETEKLYVEDEDEDEBDSLEKSDVSDSM 315  
 DB 546 -----KKLYKCNICDK--SFTWCASLTKHK--KFHTGCK-----PY 577  
 QY 316 SAQRSIKSF--TASLEGSKSVSKLISNGKKI--NCPKNCDDRMFSREYDLRRLKMH 369  
 DB 578 KCRKCGKSPQJSTLTKRHQKIHEDKHTGKLYKC--NDCDRSYSHSFRHRQKXH 632

RESULT 10  
 DB 08B185 PRELIMINARY; PRT; 634 AA.  
 AC 08B185;  
 DT 01-MAR-2003 (TEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830058L05 product:stimilar to DNA-BINDING PROTEIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Mech. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shihata Y., Hayatsu N., Sugahara Y., Shihata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shihata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Kono H., Akiyama J., Nishi K., Katsunai T., Taahiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Harana M., Nishino T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RA Adachi J., Aizawa K., Akimura T., Aikawa T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuta M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shihata K., Shingawa A., Shitaki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T., Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK043955; BACJ1714.1; -.  
 DR HSSP: P08045; 1ZNF.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro: IPR001909; KRAB.  
 DR InterPro: IPR007087; ZnF\_C2H2.  
 DR InterPro: IPR007086; ZnF\_C2H2\_sub.  
 DR Pfam: PF01352; KRAB.1.  
 DR PRINTS: PR00048; ZINCINGER.  
 DR PRODOM: PD000003; ZnF\_C2H2; 15.  
 DR SMART: SM00349; KRAB.1.  
 DR SMART: SM00355; ZnF\_C2H2; 16.  
 DR PROSITE: PS00805; KRAB.1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 16.  
 DR PROSITE: PS01517; ZINC\_FINGER\_C2H2\_2; 16.  
 KW DNA-binding.  
 SQ  
 SEQUENCE 634 AA; 73668 MW; AA7DEB33D87508BA CRC64;

Query Match 18.4%; Score 410.5; DB 2; Length 634;  
 Best Local Similarity 32.8%; Pred. No. 1,2e-15;  
 Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

QY 23 KKYCTYEGCDKAYNRPSSLEOHRLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82  
 DB 323 KPYKC--KECGKAFQTSSSLKHLNLTHTGKKPKC--EECDKSFTEKSTLTTHKRJHTGK 378  
 QY 83 KPRHCGVCGKGVNSRQHLKRHEITHT--KSFCTFENCCEAFYKHOSLR--HHILSVHEKT 139  
 DB 379 KPRKCNICGSPFNSCTNLKTHQTLHTGCKPYKC--KECGKAFYMSLSLKHQMLHNGEKP 436  
 QY 140 LTKGCKNVFTRPSKLAQHLKHHGSPAYOCDFPCFKNFQWTVLOFH--IKOSHPLK 197  
 DB 437 YKCECKDSFTLKSLSLTHQTLHTGCKP--YKCN1--CGKSFNCTNLKTHQTLHTGCKP 493  
 QY 198 KCPKCGKGVCGKGLSSHMLSHDSTWIKITWCDYCVGKFAKKNELVEHYNIFFHGNIP 257  
 DB 494 KCECKGKSFHYMSLSKSHQNLHSGE---KPYCKECD--KSFTKSTLIKQRI--HTG--- 545  
 QY 258 DDLKETEYKLE--NLDDGSKLNNLHELETEKLYVEDEDEDEBDSLEKSDVSDSM 315  
 DB 546 -----KKLYKCNICDK--SFTWCASLTKHK--KFHTGCK-----PY 577  
 QY 316 SAQRSIKSF--TASLEGSKSVSKLISNGKKI--NCPKNCDDRMFSREYDLRRLKMH 369  
 DB 578 KCRKCGKSPQJSTLTKRHQKIHEDKHTGKLYKC--NDCDRSYSHSFRHRQKXH 632

```
RESULT 11
ID 07RY67 PRELIMINARY; PRT; 611 AA.
AC 07RY67;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Hypothetical protein.
OS Name=NCU00038.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxId=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Berkeovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Seltrenihoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schlute U.,
RA Kothe G.O., Jedd C., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gierre S.,
RA Kamel M., Kamysseilis M., Mauceli E., Bielke C., Rudd S., Fishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D., Alex L.A., Mannhaupt G., Ebbole D.O., Freltag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
RL Nature 0:0-0(2003).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000729; EAA27789.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
DR KX Hypothetical protein.
SQ SEQUENCE 611 AA; 67682 MW; 995B4EAS410D8932 CRC64;

Query Match 18.3%; Score 407; DB 2; Length 611;
Best Local Similarity 24.9%; Pred. No. 1.8e-15;
Matches 111; Conservative 80; Mismatches 146; Indels 108; Gaps 17;

QY 1 MSESDETKSISSISSSSSRP-----KKYICVEGCDKADYRNPSLLBOHPRHNSD 52
DB 91 LTHADPTTVAANSITTTTRASRRPSPSLKTRCTFACCTTTPNPAALAAHLKSHND 150
QY 53 RPYKTVDDCDKAFPRKSHLETHIV-SHSEKPFHCSV--CGKGVNSRQHLKHEHTH-- 107
DB 151 RPFKCPYDDCDVLEEKLSQHIKSHNDKRYCPREPGCKSPFTNRLRLHVLVHGS 210
QY 108 TYSFKCT-FENCOEAFYKQSLRHILSVHEKTLTKCKCNKVFTPRSKLAQHLKHGGS 166
DB 211 ADHYRCRGYGDCCVQSPFRKHQTQRHRTVHL-----GK 243
QY 167 PAYQCDHPCGCFKNFQVWSVLQFIKSHPKLK--CPKC-----GKCGVKGK-----L 212
DB 244 SAYPCGNDCCDAGFDASALRRRVEHEHDLKFWCDECAEBGDDSGRRRGFTTMLL 303
QY 213 SSHMLSHDSTMIKIWTCDYCVGKPAKKNELVEHYNIPHDGNIIPDLLEKVEKLENTL 272
DB 304 GAHMKKEHN-----CAFCGV-RCGTGSNNMRHVELYHSAKYVD--RKTIACWEGC 353
QY 273 LDGSKLNLHLETEKLVKVEEDER--DEDSLDEKRSDDVRSDSMSAQRISKFTASIE 329
```

```
DB 354 DKKFTRVSNLN---THIKSAHEGHRFCGQDTDTYDAKYSIADWMPAEBCGCGFTTRVK 410
QY 330 GSKSVSKLISNSGKK-----INCPKNKCDMPFSREYDL 362
DB 411 LEEHY--LHVHKGKRPKLYVPSPMVAQAQAQAALDTRDLACPAIGCTARFTRHDL 468
QY 363 RRLHKMDNLTQRISFINSIEKEE 387
DB 469 DRHIONDHGN-----NGIKKEE 485

RESULT 12
ID 08N703 PRELIMINARY; PRT; 404 AA.
AC 08N703;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ40479.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX PubMed14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obaishi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Kato Y.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirooka S., Chida Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T., Kusano J.,
RA Kanohori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togoya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sagaki N., Aotaka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togeishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK097798; BAC05174.1; -.
DR HSSP; P08047; ISP2.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; Zf-C2H2; 14.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 14.
DR SMART; SM00355; Znf_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
SQ SEQUENCE 404 AA; 45923 MW; B1B35F62P5DF0007 CRC64;
```



Query Match	Similarity	33.9%	Pred. No. 2.9e-15;	Matches 121;	Conservative 39;	Mismatches 163;	Indels 34;	Gaps 19
QY	23	KYIICYTGCGCAKAYNRPSELLEQHLRTSHSNDPRKYCTVDDCDKAFPRKSHLLETHIVSHSEK	82					
Db	39	KPYKC--BECGCAFNRRSILLTGHKLIHNGEKRYKC--BECGGGFSSTVLTNTHKAIHAE	94					
QY	83	KPFHCISVCGKGVNSFOHLKRHEITFT--KSFKCTFENCQEAIFYKHQSLRHILSVH--EK	138					
Db	95	KPYKCEECGKASNSSSSKLMEHKRIHTGKEPKYC--BECGKAFWSMSSILTEH--KRITHAEK	151					
QY	139	TLTGQGNKVFTRPREKLAQHLKHHGGSFAYGCHDRGCFKRNQTVSVLQFH--IKQSIHPK	196					
Db	152	PYKCEECGKAFNRSSILLTKHKLIHNGEKP--KYCE--GGGKAFSKVSTLNTKAIHAEKP	208					
QY	197	LKPCPCGKCGVCKGKLSLSHMLSHDSDTWIKIWTQCYCDVGK--FAKNEIYVERHNIFFHDGN	255					
Db	209	KYCECGCAKASNSSSSKLMEHKRIH--TGKRPYKCEC--GRAFSMSSILTEHKRI--HAGE	262					
QY	256	IPDDLKETEVYKLENLIDQGSKLNNLHLETEKLVAE--DEDEDESDLDEKRDVRSDS	314					
Db	263	KP---KYCEECGCAFTWSSSFTFKHRIHAAE--KPYKCECGKGFSFTSILTGHKLIHNGE	318					
QY	315	--MSAGRIKSTSTALBEGSKSVSKLINSGKKINCPKNNCDMFEREDIRHLKMH	369					
Db	319	KRYKCECGKAF--SWSSILTEHKLIHNGEKP--BECGKAFRRSSILTGHKRIH	371					

RESULT 13

ZN43\_HUMAN STANDARD; PRT; 803 AA.

ID\_ZN43\_HUMAN STANDARD; PRT; 803 AA.

AC P17038; P28160; Q96D61;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Zinc finger protein 43 (Zinc protein HTFf) (Zinc finger protein K0X27).

DN Name=ZNF43; Synonym=K0X27, ZNF39;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=T-cell;

RX MEDLINE=91279444; PubMed=1711675;

RT "A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines."

RL Nucleic Acids Res. 19:2921-2927 (1991).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Lymph;

RX MEDLINE=2238957; PubMed=12477932; DOI=10.1073/pnae.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Alechuth S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Gibbs S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Adnan A.C., Rodriguez S., Sanchez A., Whiting R.W., Maman A., Young A.C., Shvachenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN	[3]
RP	SEQUENCE OF 38-190 FROM N.A.
RX	MEDLINE=91219421; PubMed=2023909;
RA	Bellefroid E.J., Ponclet D.A., Leccocq P.J., Revelant O.,
RA	Martial J.A.;
RT	"The evolutionarily conserved Kruppel-associated box domain defines a
RT	subfamily of eukaryotic multifingered proteins."
RL	Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
RN	[4]
RP	SEQUENCE OF 476-531 FROM N.A.
RX	TISSUE=Lymphoid;
RX	MEDLINE=91145339; PubMed=2288909;
RA	Thiesen H.-J.;
RA	"Multiple genes encoding zinc finger domains are expressed in human T
RT	cells."
RL	New Biol. 2:363-374(1990).
CC	-1- FUNCTION: May be involved in transcriptional regulation.
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC	-1- TISSUE SPECIFICITY: T and B cell lines.
CC	-1- SIMILARITY: Belongs to the kruppel C2H2-type zinc-finger protein
CC	family.
CC	-1- SIMILARITY: Contains 1 KRAB domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X59344; CAA1932.1; -
DR	EMBL; BC006528; AA016528.2; ALT_INT.
DR	EMBL; M61869; AA58674.1; -
DR	EMBL; X52358; CAA3584.1; -
DR	PIR; S26823; S26823.
DR	HSSP; P08049; 7ZNF.
DR	TRANSFAC; T04986; -
DR	Genew; HGNC:13109; ZNF43.
DR	MM; 603972; -
DR	GO; GO:0003677; F:DNA binding; TAS.
DR	InterPro; IPR001909; KRAB.
DR	InterPro; IPR007087; Znf_C2H2.
DR	InterPro; IPR007086; Znf_C2H2_gnd.
DR	Pfam; PF01352; KRAB; 1.
DR	Pfam; PF00096; zf-C2H2; 21.
DR	PRINTS; PR00048; ZINC_FINGER.
DR	Prodom; PD000003; Znf_C2H2; 22.
DR	SMART; SM00349; KRAB; 1.
DR	SMART; SM00355; Znf_C2H2; 22.
DR	PROSITE; PS50805; KRAB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
KW	DNA-binding; Metal-binding; Nuclear protein; Repeat;
KW	Transcription regulation; Zinc-finger.
FT	DOMAIN 1 69 KRAB.
FT	FT ZN_FING 168 190 C2H2-type.
FT	FT ZN_FING 196 218 C2H2-type (degenerate).
FT	FT ZN_FING 224 246 C2H2-type (degenerate).
FT	FT ZN_FING 252 274 C2H2-type (degenerate).
FT	FT ZN_FING 280 302 C2H2-type.
FT	FT ZN_FING 308 330 C2H2-type.
FT	FT ZN_FING 336 358 C2H2-type.
FT	FT ZN_FING 364 386 C2H2-type.
FT	FT ZN_FING 392 414 C2H2-type.
FT	FT ZN_FING 420 442 C2H2-type.
FT	FT ZN_FING 448 470 C2H2-type.
FT	FT ZN_FING 476 498 C2H2-type.
FT	FT ZN_FING 504 526 C2H2-type.
FT	FT ZN_FING 532 554 C2H2-type.



```

FT  ZN_PING  560  582  C2H2-type.
FT  ZN_PING  588  610  C2H2-type.
FT  ZN_PING  616  638  C2H2-type.
FT  ZN_PING  644  666  C2H2-type.
FT  ZN_PING  672  694  C2H2-type.
FT  ZN_PING  700  722  C2H2-type.
FT  ZN_PING  728  750  C2H2-type.
FT  ZN_PING  756  778  C2H2-type.
FT  ZN_PING  784  806  C2H2-type.
FT  ZN_PING  812  834  C2H2-type.
FT  ZN_PING  840  862  C2H2-type.
FT  ZN_PING  868  890  C2H2-type.
FT  ZN_PING  896  918  C2H2-type.
FT  ZN_PING  924  946  C2H2-type.
FT  ZN_PING  952  974  C2H2-type.
FT  ZN_PING  980  1002  C2H2-type.
FT  ZN_PING  1008  1030  C2H2-type.
FT  ZN_PING  1036  1058  C2H2-type.
FT  ZN_PING  1064  1086  C2H2-type.
FT  ZN_PING  1092  1114  C2H2-type.
FT  ZN_PING  1120  1142  C2H2-type.
FT  ZN_PING  1148  1170  C2H2-type.
FT  ZN_PING  1176  1198  C2H2-type.
FT  ZN_PING  1204  1226  C2H2-type.
FT  ZN_PING  1232  1254  C2H2-type.
FT  ZN_PING  1260  1282  C2H2-type.
FT  ZN_PING  1288  1310  C2H2-type.
FT  ZN_PING  1316  1338  C2H2-type.
FT  ZN_PING  1344  1366  C2H2-type.
FT  ZN_PING  1372  1394  C2H2-type.
FT  ZN_PING  1400  1422  C2H2-type.
FT  ZN_PING  1428  1450  C2H2-type.
FT  ZN_PING  1456  1478  C2H2-type.
FT  ZN_PING  1484  1506  C2H2-type.
FT  ZN_PING  1512  1534  C2H2-type.
FT  ZN_PING  1540  1562  C2H2-type.
FT  ZN_PING  1568  1590  C2H2-type.
FT  ZN_PING  1596  1618  C2H2-type.
FT  ZN_PING  1624  1646  C2H2-type.
FT  ZN_PING  1652  1674  C2H2-type.
FT  ZN_PING  1680  1702  C2H2-type.
FT  ZN_PING  1708  1730  C2H2-type.
FT  ZN_PING  1736  1758  C2H2-type.
FT  ZN_PING  1764  1786  C2H2-type.
FT  ZN_PING  1792  1814  C2H2-type.
FT  ZN_PING  1820  1842  C2H2-type.
FT  ZN_PING  1848  1870  C2H2-type.
FT  ZN_PING  1876  1898  C2H2-type.
FT  ZN_PING  1904  1926  C2H2-type.
FT  ZN_PING  1932  1954  C2H2-type.
FT  ZN_PING  1960  1982  C2H2-type.
FT  ZN_PING  1988  2010  C2H2-type.
FT  ZN_PING  2016  2038  C2H2-type.
FT  ZN_PING  2044  2066  C2H2-type.
FT  ZN_PING  2072  2094  C2H2-type.
FT  ZN_PING  2100  2122  C2H2-type.
FT  ZN_PING  2128  2150  C2H2-type.
FT  ZN_PING  2156  2178  C2H2-type.
FT  ZN_PING  2184  2206  C2H2-type.
FT  ZN_PING  2212  2234  C2H2-type.
FT  ZN_PING  2240  2262  C2H2-type.
FT  ZN_PING  2268  2290  C2H2-type.
FT  ZN_PING  2296  2318  C2H2-type.
FT  ZN_PING  2324  2346  C2H2-type.
FT  ZN_PING  2352  2374  C2H2-type.
FT  ZN_PING  2380  2402  C2H2-type.
FT  ZN_PING  2408  2430  C2H2-type.
FT  ZN_PING  2436  2458  C2H2-type.
FT  ZN_PING  2464  2486  C2H2-type.
FT  ZN_PING  2492  2514  C2H2-type.
FT  ZN_PING  2520  2542  C2H2-type.
FT  ZN_PING  2548  2570  C2H2-type.
FT  ZN_PING  2576  2598  C2H2-type.
FT  ZN_PING  2604  2626  C2H2-type.
FT  ZN_PING  2632  2654  C2H2-type.
FT  ZN_PING  2660  2682  C2H2-type.
FT  ZN_PING  2688  2710  C2H2-type.
FT  ZN_PING  2716  2738  C2H2-type.
FT  ZN_PING  2744  2766  C2H2-type.
FT  ZN_PING  2772  2794  C2H2-type.
FT  ZN_PING  2800  2822  C2H2-type.
FT  ZN_PING  2828  2850  C2H2-type.
FT  ZN_PING  2856  2878  C2H2-type.
FT  ZN_PING  2884  2906  C2H2-type.
FT  ZN_PING  2912  2934  C2H2-type.
FT  ZN_PING  2940  2962  C2H2-type.
FT  ZN_PING  2968  2990  C2H2-type.
FT  ZN_PING  2996  3018  C2H2-type.
FT  ZN_PING  3024  3046  C2H2-type.
FT  ZN_PING  3052  3074  C2H2-type.
FT  ZN_PING  3080  3102  C2H2-type.
FT  ZN_PING  3108  3130  C2H2-type.
FT  ZN_PING  3136  3158  C2H2-type.
FT  ZN_PING  3164  3186  C2H2-type.
FT  ZN_PING  3192  3214  C2H2-type.
FT  ZN_PING  3220  3242  C2H2-type.
FT  ZN_PING  3248  3270  C2H2-type.
FT  ZN_PING  3276  3298  C2H2-type.
FT  ZN_PING  3304  3326  C2H2-type.
FT  ZN_PING  3332  3354  C2H2-type.
FT  ZN_PING  3360  3382  C2H2-type.
FT  ZN_PING  3388  3410  C2H2-type.
FT  ZN_PING  3416  3438  C2H2-type.
FT  ZN_PING  3444  3466  C2H2-type.
FT  ZN_PING  3472  3494  C2H2-type.
FT  ZN_PING  3500  3522  C2H2-type.
FT  ZN_PING  3528  3550  C2H2-type.
FT  ZN_PING  3556  3578  C2H2-type.
FT  ZN_PING  3584  3606  C2H2-type.
FT  ZN_PING  3612  3634  C2H2-type.
FT  ZN_PING  3640  3662  C2H2-type.
FT  ZN_PING  3668  3690  C2H2-type.
FT  ZN_PING  3696  3718  C2H2-type.
FT  ZN_PING  3724  3746  C2H2-type.
FT  ZN_PING  3752  3774  C2H2-type.
FT  ZN_PING  3780  3802  C2H2-type.
FT  ZN_PING  3808  3830  C2H2-type.
FT  ZN_PING  3836  3858  C2H2-type.
FT  ZN_PING  3864  3886  C2H2-type.
FT  ZN_PING  3892  3914  C2H2-type.
FT  ZN_PING  3920  3942  C2H2-type.
FT  ZN_PING  3948  3970  C2H2-type.
FT  ZN_PING  3976  3998  C2H2-type.
FT  ZN_PING  4004  4026  C2H2-type.
FT  ZN_PING  4032  4054  C2H2-type.
FT  ZN_PING  4060  4082  C2H2-type.
FT  ZN_PING  4088  4104  C2H2-type.
FT  ZN_PING  4112  4128  C2H2-type.
FT  ZN_PING  4136  4152  C2H2-type.
FT  ZN_PING  4160  4176  C2H2-type.
FT  ZN_PING  4184  4200  C2H2-type.
FT  ZN_PING  4208  4224  C2H2-type.
FT  ZN_PING  4232  4248  C2H2-type.
FT  ZN_PING  4256  4272  C2H2-type.
FT  ZN_PING  4280  4296  C2H2-type.
FT  ZN_PING  4304  4320  C2H2-type.
FT  ZN_PING  4328  4344  C2H2-type.
FT  ZN_PING  4352  4368  C2H2-type.
FT  ZN_PING  4384  4400  C2H2-type.
FT  ZN_PING  4408  4424  C2H2-type.
FT  ZN_PING  4432  4448  C2H2-type.
FT  ZN_PING  4456  4472  C2H2-type.
FT  ZN_PING  4480  4496  C2H2-type.
FT  ZN_PING  4504  4520  C2H2-type.
FT  ZN_PING  4528  4544  C2H2-type.
FT  ZN_PING  4552  4568  C2H2-type.
FT  ZN_PING  4584  4600  C2H2-type.
FT  ZN_PING  4608  4624  C2H2-type.
FT  ZN_PING  4632  4648  C2H2-type.
FT  ZN_PING  4656  4672  C2H2-type.
FT  ZN_PING  4680  4696  C2H2-type.
FT  ZN_PING  4704  4720  C2H2-type.
FT  ZN_PING  4728  4744  C2H2-type.
FT  ZN_PING  4752  4768  C2H2-type.
FT  ZN_PING  4784  4800  C2H2-type.
FT  ZN_PING  4808  4824  C2H2-type.
FT  ZN_PING  4832  4848  C2H2-type.
FT  ZN_PING  4856  4872  C2H2-type.
FT  ZN_PING  4880  4896  C2H2-type.
FT  ZN_PING  4904  4920  C2H2-type.
FT  ZN_PING  4928  4944  C2H2-type.
FT  ZN_PING  4952  4968  C2H2-type.
FT  ZN_PING  4984  5000  C2H2-type.
FT  ZN_PING  5008  5024  C2H2-type.
FT  ZN_PING  5032  5048  C2H2-type.
FT  ZN_PING  5064  5080  C2H2-type.
FT  ZN_PING  5088  5104  C2H2-type.
FT  ZN_PING  5112  5128  C2H2-type.
FT  ZN_PING  5136  5152  C2H2-type.
FT  ZN_PING  5160  5176  C2H2-type.
FT  ZN_PING  5184  5200  C2H2-type.
FT  ZN_PING  5208  5224  C2H2-type.
FT  ZN_PING  5232  5248  C2H2-type.
FT  ZN_PING  5264  5280  C2H2-type.
FT  ZN_PING  5288  5304  C2H2-type.
FT  ZN_PING  5312  5328  C2H2-type.
FT  ZN_PING  5336  5352  C2H2-type.
FT  ZN_PING  5360  5376  C2H2-type.
FT  ZN_PING  5384  5400  C2H2-type.
FT  ZN_PING  5408  5424  C2H2-type.
FT  ZN_PING  5432  5448  C2H2-type.
FT  ZN_PING  5464  5480  C2H2-type.
FT  ZN_PING  5488  5504  C2H2-type.
FT  ZN_PING  5512  5528  C2H2-type.
FT  ZN_PING  5536  5552  C2H2-type.
FT  ZN_PING  5560  5576  C2H2-type.
FT  ZN_PING  5584  5600  C2H2-type.
FT  ZN_PING  5608  5624  C2H2-type.
FT  ZN_PING  5632  5648  C2H2-type.
FT  ZN_PING  5656  5672  C2H2-type.
FT  ZN_PING  5680  5696  C2H2-type.
FT  ZN_PING  5704  5720  C2H2-type.
FT  ZN_PING  5728  5744  C2H2-type.
FT  ZN_PING  5752  5768  C2H2-type.
FT  ZN_PING  5784  5800  C2H2-type.
FT  ZN_PING  5808  5824  C2H2-type.
FT  ZN_PING  5832  5848  C2H2-type.
FT  ZN_PING  5864  5880  C2H2-type.
FT  ZN_PING  5888  5904  C2H2-type.
FT  ZN_PING  5912  5928  C2H2-type.
FT  ZN_PING  5936  5952  C2H2-type.
FT  ZN_PING  5960  5976  C2H2-type.
FT  ZN_PING  5984  6000  C2H2-type.
FT  ZN_PING  6008  6024  C2H2-type.
FT  ZN_PING  6032  6048  C2H2-type.
FT  ZN_PING  6056  6072  C2H2-type.
FT  ZN_PING  6080  6096  C2H2-type.
FT  ZN_PING  6104  6120  C2H2-type.
FT  ZN_PING  6128  6144  C2H2-type.
FT  ZN_PING  6152  6168  C2H2-type.
FT  ZN_PING  6176  6192  C2H2-type.
FT  ZN_PING  6200  6216  C2H2-type.
FT  ZN_PING  6224  6240  C2H2-type.
FT  ZN_PING  6248  6264  C2H2-type.
FT  ZN_PING  6272  6288  C2H2-type.
FT  ZN_PING  6304  6320  C2H2-type.
FT  ZN_PING  6328  6344  C2H2-type.
FT  ZN_PING  6352  6368  C2H2-type.
FT  ZN_PING  6376  6392  C2H2-type.
FT  ZN_PING  6400  6416  C2H2-type.
FT  ZN_PING  6416  6432  C2H2-type.
FT  ZN_PING  6432  6448  C2H2-type.
FT  ZN_PING  6448  6464  C2H2-type.
FT  ZN_PING  6464  6480  C2H2-type.
FT  ZN_PING  6480  6496  C2H2-type.
FT  ZN_PING  6496  6512  C2H2-type.
FT  ZN_PING  6512  6528  C2H2-type.
FT  ZN_PING  6528  6544  C2H2-type.
FT  ZN_PING  6544  6560  C2H2-type.
FT  ZN_PING  6560  6576  C2H2-type.
FT  ZN_PING  6576  6592  C2H2-type.
FT  ZN_PING  6592  6608  C2H2-type.
FT  ZN_PING  6608  6624  C2H2-type.
FT  ZN_PING  6624  6640  C2H2-type.
FT  ZN_PING  6640  6656  C2H2-type.
FT  ZN_PING  6656  6672  C2H2-type.
FT  ZN_PING  6672  6688  C2H2-type.
FT  ZN_PING  6688  6704  C2H2-type.
FT  ZN_PING  6704  6720  C2H2-type.
FT  ZN_PING  6720  6736  C2H2-type.
FT  ZN_PING  6736  6752  C2H2-type.
FT  ZN_PING  6752  6768  C2H2-type.
FT  ZN_PING  6768  6784  C2H2-type.
FT  ZN_PING  6784  6800  C2H2-type.
FT  ZN_PING  6800  6816  C2H2-type.
FT  ZN_PING  6816  6832  C2H2-type.
FT  ZN_PING  6832  6848  C2H2-type.
FT  ZN_PING  6848  6864  C2H2-type.
FT  ZN_PING  6864  6880  C2H2-type.
FT  ZN_PING  6880  6896  C2H2-type.
FT  ZN_PING  6896  6912  C2H2-type.
FT  ZN_PING  6912  6928  C2H2-type.
FT  ZN_PING  6928  6944  C2H2-type.
FT  ZN_PING  6944  6960  C2H2-type.
FT  ZN_PING  6960  6976  C2H2-type.
FT  ZN_PING  6976  6992  C2H2-type.
FT  ZN_PING  6992  7008  C2H2-type.
FT  ZN_PING  7008  7024  C2H2-type.
FT  ZN_PING  7024  7040  C2H2-type.
FT  ZN_PING  7040  7056  C2H2-type.
FT  ZN_PING  7056  7072  C2H2-type.
FT  ZN_PING  7072  7088  C2H2-type.
FT  ZN_PING  7088  7104  C2H2-type.
FT  ZN_PING  7104  7120  C2H2-type.
FT  ZN_PING  7120  7136  C2H2-type.
FT  ZN_PING  7136  7152  C2H2-type.
FT  ZN_PING  7152  7168  C2H2-type.
FT  ZN_PING  7168  7184  C2H2-type.
FT  ZN_PING  7184  7200  C2H2-type.
FT  ZN_PING  7200  7216  C2H2-type.
FT  ZN_PING  7216  7232  C2H2-type.
FT  ZN_PING  7232  7248  C2H2-type.
FT  ZN_PING  7248  7264  C2H2-type.
FT  ZN_PING  7264  7280  C2H2-type.
FT  ZN_PING  7280  7296  C2H2-type.
FT  ZN_PING  7296  7312  C2H2-type.
FT  ZN_PING  7312  7328  C2H2-type.
FT  ZN_PING  7328  7344  C2H2-type.
FT  ZN_PING  7344  7360  C2H2-type.
FT  ZN_PING  7360  7376  C2H2-type.
FT  ZN_PING  7376  7392  C2H2-type.
FT  ZN_PING  7392  7408  C2H2-type.
FT  ZN_PING  7408  7424  C2H2-type.
FT  ZN_PING  7424  7440  C2H2-type.
FT  ZN_PING  7440  7456  C2H2-type.
FT  ZN_PING  7456  7472  C2H2-type.
FT  ZN_PING  7472  7488  C2H2-type.
FT  ZN_PING  7488  7504  C2H2-type.
FT  ZN_PING  7504  7520  C2H2-type.
FT  ZN_PING  7520  7536  C2H2-type.
FT  ZN_PING  7536  7552  C2H2-type.
FT  ZN_PING  7552  7568  C2H2-type.
FT  ZN_PING  7568  7584  C2H2-type.
FT  ZN_PING  7584  7600  C2H2-type.
FT  ZN_PING  7600  7616  C2H2-type.
FT  ZN_PING  7616  7632  C2H2-type.
FT  ZN_PING  7632  7648  C2H2-type.
FT  ZN_PING  7648  7664  C2H2-type.
FT  ZN_PING  7664  7680  C2H2-type.
FT  ZN_PING  7680  7696  C2H2-type.
FT  ZN_PING  7696  7712  C2H2-type.
FT  ZN_PING  7712  7728  C2H2-type.
FT  ZN_PING  7728  7744  C2H2-type.
FT  ZN_PING  7744  7760  C2H2-type.
FT  ZN_PING  7760  7776  C2H2-type.
FT  ZN_PING  7776  7792  C2H2-type.
FT  ZN_PING  7792  7808  C2H2-type.
FT  ZN_PING  7808  7824  C2H2-type.
FT  ZN_PING  7824  7840  C2H2-type.
FT  ZN_PING  7840  7856  C2H2-type.
FT  ZN_PING  7856  7872  C2H2-type.
FT  ZN_PING  7872  7888  C2H2-type.
FT  ZN_PING  7888  7904  C2H2-type.
FT  ZN_PING  7904  7920  C2H2-type.
FT  ZN_PING  7920  7936  C2H2-type.
FT  ZN_PING  7936  7952  C2H2-type.
FT  ZN_PING  7952  7968  C2H2-type.
FT  ZN_PING  7968  7984  C2H2-type.
FT  ZN_PING  7984  8000  C2H2-type.
FT  ZN_PING  8000  8016  C2H2-type.
FT  ZN_PING  8016  8032  C2H2-type.
FT  ZN_PING  8032  8048  C2H2-type.
FT  ZN_PING  8048  8064  C2H2-type.
FT  ZN_PING  8064  8080  C2H2-type.
FT  ZN_PING  8080  8096  C2H2-type.
FT  ZN_PING  8096  8112  C2H2-type.
FT  ZN_PING  8112  8128  C2H2-type.
FT  ZN_PING  8128  8144  C2H2-type.
FT  ZN_PING  8144  8160  C2H2-type.
FT  ZN_PING  8160  8176  C2H2-type.
FT  ZN_PING  8176  8192  C2H2-type.
FT  ZN_PING  8192  8208  C2H2-type.
FT  ZN_PING  8208  8224  C2H2-type.
FT  ZN_PING  8224  8240  C2H2-type.
FT  ZN_PING  8240  8256  C2H2-type.
FT  ZN_PING  8256  8272  C2H2-type.
FT  ZN_PING  8272  8288  C2H2-type.
FT  ZN_PING  8288  8304  C2H2-type.
FT  ZN_PING  8304  8320  C2H2-type.
FT  ZN_PING  8320  8336  C2H2-type.
FT  ZN_PING  8336  8352  C2H2-type.
FT  ZN_PING  8352  8368  C2H2-type.
FT  ZN_PING  8368  8384  C2H2-type.
FT  ZN_PING  8384  8400  C2H2-type.
FT  ZN_PING  8400  8416  C2H2-type.
FT  ZN_PING  8416  8432  C2H2-type.
FT  ZN_PING  8432  8448  C2H2-type.
FT  ZN_PING  8448  8464  C2H2-type.
FT  ZN_PING  8464  8480  C2H2-type.
FT  ZN_PING  8480  8496  C2H2-type.
FT  ZN_PING  8496  8512  C2H2-type.
FT  ZN_PING  8512  8528  C2H2-type.
FT  ZN_PING  8528  8544  C2H2-type.
FT  ZN_PING  8544  8560  C2H2-type.
FT  ZN_PING  8560  8576  C2H2-type.
FT  ZN_PING  8576  8592  C2H2-type.
FT  ZN_PING  8592  8608  C2H2-type.
FT  ZN_PING  8608  8624  C2H2-type.
FT  ZN_PING  8624  8640  C2H2-type.
FT  ZN_PING  8640  8656  C2H2-type.
FT  ZN_PING  8656  8672  C2H2-type.
FT  ZN_PING  8672  8688  C2H2-type.
FT  ZN_PING  8688  8704  C2H2-type.
FT  ZN_PING  8704  8720  C2H2-type.
FT  ZN_PING  8720  8736  C2H2-type.
FT  ZN_PING  8736  8752  C2H2-type.
FT  ZN_PING  8752  8768  C2H2-type.
FT  ZN_PING  8768  8784  C2H2-type.
FT  ZN_PING  8784  8800  C2H2-type.
FT  ZN_PING  8800  8816  C2H2-type.
FT  ZN_PING  8816  8832  C2H2-type.
FT  ZN_PING  8832  8848  C2H2-type.
FT  ZN_PING  8848  8864  C2H2-type.
FT  ZN_PING  8864  8880  C2H2-type.
FT  ZN_PING  8880  8896  C2H2-type.
FT  ZN_PING  8896  8912  C2H2-type.
FT  ZN_PING  8912  8928  C2H2-type.
FT  ZN_PING  8928  8944  C2H2-type.
FT  ZN_PING  8944  8960  C2H2-type.
FT  ZN_PING  8960  8976  C2H2-type.
FT  ZN_PING  8976  8992  C2H2-type.
FT  ZN_PING  8992  9008  C2H2-type.
FT  ZN_PING  9008  9024  C2H2-type.
FT  ZN_PING  9024  9040  C2H2-type.
FT  ZN_PING  9040  9056  C2H2-type.
FT  ZN_PING  9056  9072  C2H2-type.
FT  ZN_PING  9072  9088  C2H2-type.
FT  ZN_PING  9088  9104  C2H2-type.
FT  ZN_PING  9104  9120  C2H2-type.
FT  ZN_PING  9120  9136  C2H2-type.
FT  ZN_PING  9136  9152  C2H2-type.
FT  ZN_PING  9152  9168  C2H2-type.
FT  ZN_PING  9168  9184  C2H2-type.
FT  ZN_PING  9184  9200  C2H2-type.
FT  ZN_PING  9200  9216  C2H2-type.
FT  ZN_PING  9216  9232  C2H2-type.
FT  ZN_PING  9232  9248  C2H2-type.
FT  ZN_PING  9248  9264  C2H2-type.
FT  ZN_PING  9264  9280  C2H2-type.
FT  ZN_PING  9280  9296  C2H2-type.
FT  ZN_PING  9296  9312  C2H2-type.
FT  ZN_PING  9312  9328  C2H2-type.
FT  ZN_PING  9328  9344  C2H2-type.
FT  ZN_PING  9344  9360  C2H2-type.
FT  ZN_PING  9360  9376  C2H2-type.
FT  ZN_PING  9376  9392  C2H2-type.
FT  ZN_PING  9392  9408  C2H2-type.
FT  ZN_PING  9408  9424  C2H2-type.
FT  ZN_PING  9424  9440  C2H2-type.
FT  ZN_PING  9440  9456  C2H2-type.
FT  ZN_PING  9456  9472  C2H2-type.
FT  ZN_PING  9472  9488  C2H2-type.
FT  ZN_PING  9488  9504  C2H2-type.
FT  ZN_PING  9504  9520  C2H2-type.
FT  ZN_PING  9520  9536  C2H2-type.
FT  ZN_PING  9536  9552  C2H2-type.
FT  ZN_PING  9552  9568  C2H2-type.
FT  ZN_PING  9568  9584  C2H2-type.
FT  ZN_PING  9584  9600  C2H2-type.
FT  ZN_PING  9600  9616  C2H2-type.
FT  ZN_PING  9616  9632  C2H2-type.
FT  ZN_PING  9632  9648  C2H2-type.
FT  ZN_PING  9648  9664  C2H2-type.
FT  ZN_PING  9664  9680  C2H2-type.
FT  ZN_PING  9680  9696  C2H2-type.
FT  ZN_PING  9696  9712  C2H2-type.
FT  ZN_PING  9712  9728  C2H2-type.
FT  ZN_PING  9728  9744  C2H2-type.
FT  ZN_PING  9744  9760  C2H2-type.
FT  ZN_PING  9760  9776  C2H2-type.
FT  ZN_PING  9776  9792  C2H2-type.
FT  ZN_PING  9792  9808  C2H2-type.
FT  ZN_PING  9808  9824  C2H2-type.
FT  ZN_PING  9824  9840  C2H2-type.
FT  ZN_PING  9840  9856  C2H2-type.
FT  ZN_PING  9856  9872  C2H2-type.
FT  ZN_PING  9872  9888  C2H2-type.
FT  ZN_PING  9888  9904  C2H2-type.
FT  ZN_PING  9904  9920  C2H2-type.
FT  ZN_PING  9920  9936  C2H2-type.
FT  ZN_PING  9936  9952  C2H2-type.
FT  ZN_PING  9952  9968  C2H2-type.
FT  ZN_PING  9968  9984  C2H2-type.
FT  ZN_PING  9984  10000  C2H2-type.

```

```

CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: AB040906; BAA59597.1; -.
CC  HSSP: P07248; IARD.
CC  Genew: HGNC:23707; ZNF492.
CC  InterPro: IPR001909; KRAB.
CC  InterPro: IPR007087; Znf_C2H2.
CC  InterPro: IPR007086; Znf_C2H2_sub.
CC  Pfam: PF01352; KRAB; 1.
CC  Pfam: PF00096; zf-C2H2; 13.
CC  PRINTS: PR00048; ZINC_FINGER.
CC  ProDom: PD000003; Znf_C2H2; 12.
CC  ProSite: PS50805; KRAB; 1.
CC  ProSite: PS50028; ZINC_FINGER_C2H2_1; 12.
CC  ProSite: PS50157; ZINC_FINGER_C2H2_2; 13.
CC  DNA-binding: Metal-binding; Nuclear protein; Repeat;
CC  Transcription regulation; Zinc-finger.
CC  NON_TER 1 1
CC  DOMAIN 15 86
CC  ZN_FING 184 206 KRAB.
CC  ZN_FING 212 234 C2H2-type 1.
CC  ZN_FING 240 262 C2H2-type 2.
CC  ZN_FING 268 290 C2H2-type 3.
CC  ZN_FING 296 318 C2H2-type 4.
CC  ZN_FING 324 346 C2H2-type 5.
CC  ZN_FING 352 374 C2H2-type 6.
CC  ZN_FING 380 402 C2H2-type 
```

## RESULT 15

```

ID      O43693      PRELIMINARY;      PRT;      395 AA.
AC      O43693;
DT      01-JUN-1998 (TrEMBLrel. 06, Created)
DT      01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Zinc-finger protein (Fragment).
GN      Name=ZFS-1;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Seminoma;
RX      MEDLINE=98069676; PubMed=9406578; DOI=10.1016/S0165-4608(97)00004-6;
RA      Ogawa T., Poncellet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,
RA      Udagawa K., Iecocq P., Martine J., Martial J., Hosaka M.;
RT      "Enhanced expression in seminoma of human zinc finger genes located on
RT      chromosome 19."
RL      Cancer Genet. Cytogenet. 100:36-42(1998).
DR      EMBL; D70831; BAA2450.1; -.
DR      HSSP; P08153; 1ZFD.
DR      GO; GO:0005634; C:nucleus; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      GO; GO:0008270; F:zinc ion binding; IEA.
DR      InterPro; IPR006025; Pept_M_Zn_BS.
DR      InterPro; IPR007087; ZnF_C2H2.
DR      InterPro; IPR007086; ZnF_C2H2_sub.
DR      Pfam; PF00096; zf-C2H2; 13.
DR      ProDom; PD000048; ZINC_FINGER.
DR      PRINTS; PR000003; ZnF_C2H2; 12.
DR      ProDom; PD000003; ZnF_C2H2; 12.
DR      SMART; SM00355; ZnF_C2H2; 13.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR      PROSITE; PS00157; ZINC_FINGER_C2H2_2; 13.
DR      PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW      Zinc-finger.
KW      ZINC_FINGER.
FT      NON_TER      1      1
SQ      SEQUENCE      395 AA; 45291 MW; D4E06B54DCE0BF9B CRC64;

Query Match      17.8%; Score 397.5; DB 2; Length 395;
Best Local Similarity 35.9%; Pred. No. 4e-15;
Matches 97; Conservative 35; Mismatches 87; Indels 51; Gaps 14;

QY      23 KXYICTYEGCDKAYNRPSSLLQHLRTHSNDPRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82
DB      133 KPYKC--KECGKAFNRSSLTLTRKIHTEGKPYKC--EECGKALQSSHLTRKIHTEG 188
QY      83 KPRHSCVCCGKGVNSRQHLKRNHEITHT--KSFKCTFENCQEA----- 122
DB      189 KPYKCKKCCGAFQSAHLTTHEVTHTEGKPYKC--EKCGKAFNHPSHLTTHKLIHTEGKP 246
QY      123 -----YKHOS--LRHHILSVHEKTLCTKQCNKVFTRPSKLAQHKLKHGGSPAYOC 171
DB      247 YKCEGCKAFKHSSTLTRKLIHTEGKPYKCEGKAFQSSKLTENK--KIHTGEKPYEC 305
QY      172 DHPGCFKNFQTVSWLQPHIKQSHPK--LKCPRCKGKGVGKGLSSHMLSHDSTWIKIW 228
DB      306 EE--CGKAFQSSNLTGRH--KKSHTGEKPYKCEGKALNRPSTLTIHKIHIH---TGEKPY 359
QY      229 TCDYCDVGK--FAKKNELVEHYNI FHDGNIP 257
DB      360 KCEEC--GKAFQSSKLTTRKHI--HTGEKPY 386
```

Search completed: February 8, 2005, 15:22:03  
Job time : 183 secs

This Page Blank (uspto)